Chaudhuri, Ananish (Ed.)

Book — Published Version

Recent advances in experimental studies of social dilemma games

Suggested Citation: Chaudhuri, Ananish (Ed.) (2016) : Recent advances in experimental studies of social dilemma games, ISBN 978-3-03842-206-8, MDPI, Basel, https://doi.org/10.3390/books978-3-03842-206-8

games

Recent Advances in Experimental Studies of Social Dilemma Games

Edited by
Ananish Chaudhuri

Printed Edition of the Special Issue Published in *Games*

www.mdpi.com/journal/games

MDPI

Ananish Chaudhuri (Ed.)

Recent Advances in Experimental Studies of Social Dilemma Games

To Ananrita, Ishannita and Indira

This book is a reprint of the Special Issue that appeared in the online, open access journal, *Games* (ISSN 2073-4336) from 2015–2016 (available at: http://www.mdpi.com/journal/games/special_issues/experimental-social-dilemma).

Guest Editor
Ananish Chaudhuri
University of Auckland
New Zealand

Editorial Office
MDPI AG
Klybeckstrasse 64
Basel, Switzerland

Publisher
Shu-Kun Lin

Assistant Editor
Haiqing Chen

1. Edition 2016

MDPI • Basel • Beijing • Wuhan • Barcelona

ISBN 978-3-03842-231-0 (Hbk)
ISBN 978-3-03842-206-8 (PDF)

Table of Contents

Section 3

List of Contributors

Priyodorshi Banerjee Economic Research Unit, Indian Statistical Institute, Kolkata 700108, India.

Ralph-C. Bayer School of Economics, University of Adelaide, Adelaide SA 5005, Australia.

Eva Camacho-Cuena Department of Economics, University Jaume I, Castellón 12071, Spain.

Juan Camilo Cárdenas Economics Department, Universidad de los Andes, Bogotá 111711, Colombia.

Sujoy Chakravarty Centre for Economic Studies and Planning, School of Social Sciences, Jawaharlal Nehru University, New Delhi 110067, India.

Ananish Chaudhuri Department of Economics, University of Auckland, 660 Owen G Glenn Building Level 6, 12 Grafton Road, Auckland 1142, New Zealand.

Brice Corgnet Argyros School of Business and Economics, Chapman University, 1 University Drive, Orange, CA 92866, USA; Economic Science Institute, Chapman University, 1 University Drive, Orange, CA 92866, USA.

Caleb A. Cox Department of Economics, Virginia Commonwealth University, Snead Hall, 301 W. Main Street, P.O. Box 844000, Richmond, VA 23284-4000, USA.

Lachlan Deer Department of Economics, University of Zurich, Schönberggasse 1, Zurich CH 8001, Switzerland.

Davide Dragone Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy.

Michalis Drouvelis Department of Economics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

Fabio Galeotti Groupe d'Analyse et de Théorie Economique Lyon St. Etienne (GATE-LSE), CNRS and University of Lyon, 93, Chemin des Mouilles, 69130 Ecully, France.

Sanmitra Ghosh Department of Economics, Jadavpur University, Kolkata 700032, India.

Ernan Haruvy Jindal School of Management, The University of Texas at Dallas, SM 32, 800 West Campbell Road, Richardson, TX 75080, USA.

Karen Evelyn Hauge Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, Oslo 0349, Norway.

Roberto Hernán-González Business School, Nottingham University, Jubilee Campus, Nottingham, NG8 1BB, UK.

Yoshio Iida Department of Economics, Kyoto Sangyo University, Kyoto 6038555, Japan.

César Mantilla Institute for Advanced Study in Toulouse, Toulouse School of Economics, Toulouse 31015, France.

Matthew W. McCarter College of Business, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249, USA; Economic Science Institute, Chapman University, 1 University Drive, Orange, CA 92866, USA.

Raimondello Orsini Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy.

Peter T. L. Popkowski Leszczyc School of Business, University of Alberta, Edmonton, AB T6G-2R6, Canada.

Till Requate Department of Economics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany.

Ole Rogeberg Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, Oslo 0349, Norway.

Rajiv Sethi Department of Economics, Barnard College, Columbia University, New York, NY 10027, USA.

Aric P. Shafran Department of Economics, California Polytechnic State University, San Luis Obispo, CA 93407, USA.

Shosh Shahrabani Department of Economics, The Max Stern Yezreel Valley College, 19300 Emek Yezreel, Israel.

Ch'ng Kean Siang Department of Economics, School of Social Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia.

Brock Stoddard Department of Economics, University of South Dakota, Beacom Hall, 414 E Clark St., Vermillion,SD 57069, USA.

Israel Waichman Department of Economics, University of Heidelberg, Berheimerstr. 20, Heidelberg 69115, Germany.

About the Guest Editor

Ananish Chaudhuri is the Professor of Experimental Economics and Head of the Department of Economics at the University of Auckland. His research has appeared in leading scholarly journals such as *Review of Economic Studies, Economic Journal, Journal of Public Economics* and *Experimental Economics*. He is the author of the critically acclaimed book "*Experiments in Economics: Playing Fair with Money*", published by Routledge. The book provides an easy to follow guide to economic experiments and specifically those that explore notions of fairness, altruism and trust in economic transactions and how findings in the field can change the way we approach a variety of economic problems such as pricing by firms, written contracts between parties, making voluntary contributions to charity or the provision of micro-credit to small entrepreneurs. Between 2007 and 2010, he served as the Editor-in-Chief of the *New Zealand Economic Papers*, which is published by Taylor and Francis, on behalf of the New Zealand Association of Economists. He currently serves as an Associate Editor and the Book Review Editor for the *Journal of Economic Psychology* and is an Associate Editor of the *Journal of Behavioral and Experimental Economics*. Between 2011 and 2015, he also served as an Associate Editor for the *Journal of Economic Behavior and Organization*. He hails from Calcutta, India and holds a B.Sc. in Economics from Presidency College Calcutta, an M.A. in Economics from Jawaharlal Nehru University New Delhi and a Ph.D. in Economics from Rutgers University, New Brunswick, New Jersey.

Preface to "Recent Advances in Experimental Studies of Social Dilemma Games"

Ananish Chaudhuri

Abstract: I provide a broad overview of the findings reported in the articles submitted for this special volume on experimental studies of social dilemma problems. I start by providing a synopsis of where current research stands on this topic. Then I go on to discuss the specific papers and how those papers extend our knowledge in this area and add value to the current state of the art.

Reprinted from *Games*. Cite as: Chaudhuri, A. Recent Advances in Experimental Studies of Social Dilemma Games. *Games* **2016**, *7*, 7.

1. Introduction

A wide range of situations in life can be thought of as social dilemmas which pose tension between cooperation and self-interest. Such situations range from voting; to contributions to charities and public goods; to managing common pool resources; to even tackling global warming and climate change. In all these situations individuals face a trade-off: cooperation makes everyone better off, but as long as enough others cooperate in order to achieve the goal, then individual self-interest suggests free-riding on the cooperation of others. If enough participants give in to this inherent incentive for free-riding then sustaining cooperation is infeasible.[1]

---

[1] Joseph Heller captures the logic behind free-riding eloquently in the following passages from his book Catch-22.

> *Sharing a tent with a man who was crazy wasn't easy but Nately didn't care. He was crazy, too, and had gone every free day to work on the officers' club that Yossarian had not helped build . . . .Actually, there were many officers' clubs that Yossarian had not helped build, but he was the proudest of the one on Pianosa. It was a sturdy and complex monument to his powers of determination. Yossarian never went there to help until it was finished; then he went there often, so pleased was he with the large, fine, rambling shingled building. It was truly a splendid structure, and Yossarian throbbed with a mighty sense of accomplishment each time he gazed at it and reflected that none of the work that had gone into it was his.*

A little later in the book, Heller goes on to explain how this line of thinking leads to the unravelling of cooperation. This is represented in the following conversation between the book's protagonist Yossarian and Major Major Major Major, whose first, middle and last names are Major and who also holds the rank of major.

Researchers in the area have tended to focus on two types of social dilemma games. One of these is the voluntary contributions mechanism. Here, members of a group are provided with an initial endowment and have the choice of contributing to a public good that generates benefits for all group members regardless of whether they contributed or not. The social optimum in these games is for all members to contribute their entire endowment to the public good with the attendant benefits being distributed equally among all group members. The self-interested or individually rational course of action, of course, is to contribute nothing and free-ride on others' contribution. Free-riding is the dominant strategy Nash equilibrium in one-shot plays of the game as well as the subgame perfect outcome in repeated plays.

The second version of a social dilemma studied by researchers is the common pool resource extraction problem which simulates harvesting common resources such as logging or fishing. Here, the social optimum is for every player to abide by their pre-assigned quota for harvesting, while the self-interested action is to engage in over-extraction of the resource. In the Nash equilibrium of this game, everyone extracts more than their quota and the resource is depleted. Given the prevalence of such dilemmas in most spheres of human interaction, the problem has drawn the attention of researchers not only in economics but across different disciplines in the social sciences. Bohm (1972, 1983) [1,2] and Marwell and Ames (1979, 1980, 1981) [3–5] undertook some of the earliest experimental work in this area. Notable early contributions from economists include Andreoni (1988, 1990, 1995) [6–8], Isaac and Walker (1988a, 1988b) [9,10], and Isaac, McCue and Plott (1985) [11]. Noteworthy early studies from other social sciences include Dawes, McTavish and Shaklee (1977) [12], Dawes (1980) [13], Dawes, Orbell, Simmons and van de Kragt (1986) [14], Ostrom (1990) [15], Ostrom, Gardner and Walker (1994) [16], Ostrom, Walker and Gardner (1992) [17] and Yamagishi (1986, 1988) [18,19].

It is, by now, well-documented across the social sciences that humans are willing to engage in far-more cooperation than the self-interested model would suggest. The strong Nash equilibrium prediction of complete free-riding or over-extraction is seldom borne out with typically much greater levels of cooperation than predicted.

"Suppose we let you pick your missions and fly milk runs," Major Major said. "That way you can fly the four missions and not run any risks."

"I don't want to fly milk runs. I don't want to be in the war anymore."

"Would you like to see our country lose?" Major Major asked.

"We won't lose. We've got more men, more money and more material. There are ten million people in uniform who can replace me. Some people are getting killed and a lot more are making money and having fun. Let somebody else get killed."

"But suppose everybody on our side felt that way."

"Then I'd certainly be a damned fool to feel any other way. Wouldn't I?"

Such cooperation manifests itself in at least two ways. First, humans cooperate more than expected even in one-off interactions, which rules out the possibility of future benefits to co-operative behavior. In fact, a plurality, if not a majority of subjects are conditional cooperators whose degree of co-operation is positively correlated with their beliefs about their peers. Those who expect their peers to co-operate do so; while those who are pessimistic about their peers' level of co-operation end up free-riding.

Second, not only are humans willing to cooperate, they are also willing to engage in "altruistic punishment" in the sense that they are willing to forego money to inflict punishment on violators of cooperative norms; not only in repeated plays but even in one-off interactions with no possibility of such punishments yielding future benefits. This tendency has been labelled as "strong reciprocity". See, for instance, Fehr and Gächter (2000, 2002) [20,21], Gintis *et al.* (2005) [22] and Yamagishi (1986, 1988) [18,19]. In fact socio-biological explanations of human cooperation such as Hamilton's (1964) [23] theory of kin selection or Trivers's (1971) [24] theory of reciprocal altruism or the concept of an evolutionary stable strategy (Maynard Smith, 1982) [25] are not able to explain large patterns of human cooperation. This is, at least partly, because they ignore the influence of culture or social norms on human behavior. Emotional dispositions such as fairness, trust and reciprocity often contribute to creating and sustaining virtuous social norms that can be passed on in the history of human societies. (Axelrod, 1986 [26]).

However, while it is certainly true that the level of cooperation is typically well above that predicted by a model based on individual self-interest, it is also the case that over time, as the interaction is repeated, there is a tendency for agents to reduce their level of cooperation. However, even with this decay the strong Nash equilibrium prediction of complete free riding (in the case of the public goods game) or maximal extraction (in the case of the common pool resource game), is seldom borne out. This, in turn, has led to two separate research questions. The first question is about the nature of the game itself; how it is perceived by those playing it; what strategies do the players adopt and what leads to the decay in contributions over time. The second research agenda seeks to identify mechanisms or institutions that can sustain cooperation over time.

Early writing in experimental economics identified a number of different factors that might cause this pattern of decay. These included kindness on the part of some and confusion on the part of others (Andreoni, 1995 [6]), the "warm glow" of giving (Andreoni, 1990 [7]), a combination of learning to play the dominant strategy and strategic play by self-interested players (Andreoni, 1988 [8]) and decision errors of various types (Palfrey and Prisbrey, 1997 [27]; Anderson, Goeree and Holt, 1998 [28]). Recent research, however, has honed in on two possible explanations for declining contributions. Both of these rely on the presence of "*conditional cooperators*",

whose contributions are positively correlated to their beliefs about the contributions of other group members (see Fischbacher *et al.*, 2001 [29]). One strand of this literature, and indeed the broader literature on social dilemma games, assumes that any population is composed of at least two types of players: cooperators and free-riders. Cooperators start out making high contributions, but over time recognize the presence of free-riders, and reduce their contributions in retaliation, leading to the oft-seen pattern of decay. See Gunnthorsdottir, Houser and McCabe (2007) [30] for an example of this kind of argument.

Ambrus and Pathak (2011) [31] provide a formal theoretical model along these lines. They assume that there are two kinds of players: purely self-interested and reciprocal. The proportion of each type is common knowledge, so there is no asymmetric information about types. The main feature of the model is a reciprocity function with arguments including past and current contributions of other group members. The authors show that, if the reciprocity function obeys certain regularity conditions, then there is a unique sub-game perfect equilibrium in which contributions decline over time. In sum, this line of literature assumes that populations consist of people with different preference types.

A recent study by Chaudhuri, Paichayontvijit and Smith (2015) [32] argues that while the presence of different preference types may be sufficient for contributions to decay, this may not be necessary. These authors show that even if the majority of players are conditional co-operators with few or no free-riders, as long as there is enough heterogeneity in the prior beliefs of those conditional co-operators, contributions will still decay. This is primarily due to the fact that players adjust their contributions to converge to the group average; but those who are above the average reduce their contributions by a greater magnitude relative to those who are below the average and this inevitably leads to decaying contributions.

The other strand of the literature is based on the idea of self-serving biases. Fischbacher and Gächter (2010) [33], for example, report that a *"self-serving bias"* in conditional cooperation, where each agent attempts to contribute slightly less than the group average, leads to the contribution decay. Neugebauer, Perote, Schmidt, and Loos (2009) [34] provide further evidence on self-serving biases. They examine subjects' own contributions and their beliefs about others' contributions when subjects get feedback about others' contributions and when they do not. Contributions decline only when feedback is provided, suggesting that behavior is adaptive when subjects get feedback. The authors report that contributions and beliefs are higher in the treatment without feedback than the one with feedback and that contributions are positively correlated with beliefs. Smith (2013) [35] extends this line of work by addressing the issue of beliefs being endogenous to contribution decisions. Using an instrumental variables approach, he estimates the causal effect of beliefs on contributions.

Given the pattern of gradual decay in cooperation levels, a second research question tries to understand what institutions or mechanisms may prevent such decay and sustain virtuous norms of cooperation over longer horizons. Axelrod (1986) [26] suggests that a social norm is essentially an implicit rule that members of society feel compelled to adhere to. One way of creating and sustaining such a norm is via internalization, where a norm becomes so entrenched in a society that violating it causes psychological discomfort. Such internalization may be achieved via punishments of norm-violators, where said punishment can be monetary or may take non-monetary forms such as expressions of disapproval, or social ostracism and exclusion. Cooperation may also be achieved via devices that can be broadly described as *moral suasion* including exhortative advice from prior groups of players, promises of cooperation, generating feelings of community etc. Researchers have also looked at the efficacy of communication combined with the ability to punish norm violators; both in the case of groups that are formed exogenously (determined by the experimenter) or endogenously (where group members can choose who they wish to interact with). Not surprisingly, given the wide-spread appeal and interest in such topics across the social-sciences, the literature looking at aspects of this is voluminous. Ledyard (1995) [36] provides a comprehensive overview from the initial days of this work till *circa* 1995 while Chaudhuri (2011) [37] provides an update on this line of work since Ledyard (1995) [36] till around 2010.

The papers in this volume make significant contributions to our understanding of some of these questions and thereby help advance the frontiers of knowledge in this area. It is possible to informally divide the ten papers contained in this volume into three sections, with a unifying theme for each of the three. The three papers in the first section address the first research stream highlighted above; they all explore structural issues of social dilemma games, in the sense that they ask fundamental questions about the parameters of the game, subject perceptions and the strategies adopted when engaging in such social dilemmas. The next three papers all explore the second of the two above-mentioned research questions, in that they ask how we may be able to sustain cooperation in social dilemmas over the longer term. The final cluster of four papers can be thought of as innovative applications of the social dilemma paradigm to understanding a number of real-world phenomena. In Section 2 below, I provide an overview of the papers in each of those three clusters. I provide some concluding remarks in Section 3.

2. Overview of the Studies in This Volume

2.1. Papers in Section 1

In the first paper, Caleb Cox and Brock Stoddard attempt to resolve a long-standing controversy as to how levels of cooperation are affected by the

framing of the underlying game and the nature of feedback provided. Received wisdom suggests that when the underlying game is framed as one involving resource extraction then we should expect less cooperation and greater free-riding (in the form of over-extraction) as opposed to a situation where the game is framed as involving contributions to a public good; in the latter case, we would expect comparatively greater levels of cooperation. Cox and Stoddard construct payoff-equivalent games where in the "give" frame subjects can choose to contribute to a public good, while in the "take" frame they choose to extract from a common pool. The authors also vary the matching mechanism, whether group membership either remains unchanged (Partners) or members are randomly re-matched from one round to the next (Strangers) as well as the nature of the feedback provided; whether subjects receive feedback only at the aggregate level or whether they can see individual decisions. Contrary to prior findings, it is not the case that, on average, there is greater cooperation in the "give" frame; if anything, on average, there is greater cooperation in the "take" frame (particularly with individual level feedback) or no strong differences. It is also not the case that individual level feedback leads to greater free-riding. However, it is certainly the case that using a "take" framing and providing more fine-grained feedback at the individual-level leads to greater variability in co-operation and more extreme behavior (in terms of both greater free-riding and full cooperation), especially when groups are fixed (Partners matching).

In the second paper of this section, Juan Camilo Cardenas, Cesar Mantilla and Rajiv Sethi provide an alternative way of analyzing how subjects make decisions in a common pool resource extraction game. As noted above, behavior of subjects in such experiments typically deviates substantially from the equilibrium prediction in that all available actions are chosen with a positive frequency, with strictly dominated actions (less than maximal extraction) being chosen persistently and often. Average extraction is relatively stable over time, lying below equilibrium levels, but above efficient levels. The authors rely on data collected in the two common pool resource extraction studies undertaken by Ostrom *et al.* (1992) [17] and Cardenas (2004) [38] and argue that these patterns can be accurately replicated with a model of payoff sampling equilibrium developed by Osborne and Rubinstein (1998) [39] with some suitable refinements.

The basic idea underlying this solution concept is that individuals try out multiple actions, observe payoffs and subsequently adopt actions that were the most rewarding. A sampling equilibrium is a distribution of actions in a population that reproduces itself, in the sense that the likelihood with which an action is selected under the sampling procedure matches the frequency with which it is currently being used. The authors compare the sampling equilibrium with the more widely-used solution concept of quantal response equilibrium (QRE) introduced by McKelvey

and Palfrey (1995) [40] , which is based on the idea that individuals make errors when responding to the behavior of others, have accurate beliefs about the distribution of opponent actions and best respond to these beliefs. The authors focus their attention on the logit QRE model, which has one free parameter to capture the rate at which errors are made. The logit QRE explains data from such games extremely well but that does require fine-tuning the free parameter to fit the specific study and its value can vary widely across treatments, even within the class of common pool resource games. If the QRE parameter is constrained to be equal across different games, then payoff sampling provides a superior fit to the data. Cardenas *et al.* argue that their findings are likely to generalize to other complex games with multiple players and strategies. This then provides a road-map to alternative empirical means of analyzing decisions in these games and can help enrich our understanding of behavior in the same.

In third and final paper of this section, Karen Evelyn Hauge and Ole Rogeberg adopt a unique perspective on the public goods game. In a typical public goods game decisions are made by subjects whose payoffs are directly affected by those same decisions. Yet, Hauge and Rogeberg point out, in a variety of social dilemma situations, the decision-making power and authority is delegated to representatives who make binding decisions on behalf of a larger group. In this paper, the authors compare contribution decisions made by individuals with contribution decisions made by group representatives. Their main finding is that contribution behavior differs between individuals and group representatives, but only for women. While men's choices are equally self-interested as individuals and group representatives, women make more cooperative choices as group representatives. This has implications about deciding who should be selected to represent group interests depending on the specific context and the goals intended to be achieved.

2.2. Papers in Section 2

There are three papers in this section, all of which focus on the second of the two research questions identified above; viz., they all focus on the issue of sustaining cooperative behavior among group members in social dilemma games over the longer term. The first paper in this section is an innovative study that could also have featured in the final section of this volume as an ingenious application of the social dilemma paradigm. In this paper, Brice Corgnet, Roberto Gonzalez and Matthew McCarter design an intriguing experiment to study the problem of cyber-loafing. Cyber-loafing is a major problem facing organizations where workers, instead of concentrating on the task at hand, spend their time surfing the internet, resulting in loss of workgroup productivity. The authors seek to understand how changes in the decision-making rights about what workgroup members can do on the job affect cyber-loafing and subsequent work productivity. Subjects in this experiment

are asked to undertake a monotonous number adding task designed to simulate a data-entry organization. Each member completing a task generates payoff for the group; the social optimum is for each member to complete as many tasks as possible but if some members are working, thereby generating revenue for the group, then others can free-ride by engaging in cyber-loafing. This is what turns this data-entry game into a social dilemma. The authors compare two different types of decision-making regimes: autocratic decision-making, where an experimental monitor ("the boss") decides to turn off internet access part way through the session, while in group voting, group members vote whether to retain or turn off internet access. They report that, while autocratic decision-making and group voting regimes both curtail cyber-loafing (by over 50%), it is only in group voting that there is a substantive improvement (of 38%) in a cyber-loafer's subsequent work performance. Unlike autocratic decision-making, group voting leads to workgroups outperforming the control condition where cyber-loafing could not be stopped. Additionally, only in the group voting regime did production levels of cyber-loafers and non-loafers converge over time.

In the second paper of this section, Lachlan Deer and Ralph Bayer look at the role of pledges of commitment in a public goods game. There is a voluminous literature on this topic but what Deer and Bayer add to the literature is that they look at the efficacy of pledges where subjects can enter into partnerships but dissolving existing relationships subsequently is costly and the costs are distributed symmetrically or asymmetrically across partners. As the authors point out, there are a number of situations in life where dissolving partnerships is costly. In the marriage market individuals learn about the cooperativeness of their partners in an initial dating stage and may decide to pledge commitment through marriage. Dissolving a marriage through divorce is often costly and the size and distribution of these costs may fall differently on the two partners. Business partnerships can also be thought of as being analogous to a social dilemma with costly dissolution, where partners choose between working to enhance the value of the company versus working for oneself or shirking. Dissolving these partnerships may either be cost-free via informal agreements or costly via lost investments and/or costs to retrieve those investments.

Deer and Bayer find, in line with prior studies, that pledges of commitment alone can increase cooperation and welfare in committed partnerships. The introduction of relatively large and equally split costs yields similar gains. In contrast, pledges of commitment fail to improve cooperation and welfare when the costs to dissolve committed partnerships fall solely on the individual choosing to break up. This result is interesting since much of the earlier work demonstrating the effectiveness of promises or other similar communication mechanisms look at situations where dissolving groups is either costless or shared equally among group members. Deer and Bayer show that in the presence of asymmetric dissolution costs there may be

limits to the efficacy of promise-making. When the costs for partnership dissolution are asymmetric and fall solely on the instigator of the breakup, cooperation levels decline. This is mostly due to the fact that in the presence of asymmetric costs, the subject who does not bear that cost anticipates the reluctance of the partner (who bears the cost) to dissolve the relationship; this, in turn, causes the former to act in an opportunistic manner by free-riding.

In the third paper of this section, Michalis Drouvelis revisits the rich literature on altruistic punishments as in Fehr and Gächter (2000, 2002) [20,21]. Except Drouvelis extends current work by comparing behavior in two distinct punishment regimes: first, a standard public goods game with punishment in which subjects are given the opportunity to punish other group members (democratic punishment regime) and second, a public goods game environment where all group members exogenously experience an automatic reduction of their income, irrespective of their behavior (undemocratic punishment regime). However, in the second treatment, subjects do have the power to forego some money to alleviate any punishment for non-free-riders who have been punished unjustly. Drouvelis employs a within-subjects design where subjects experience both environments, democratic and undemocratic punishments. The design is counter-balanced to control for order effects. His findings indicate that average contributions and earnings in the undemocratic punishment environment are significantly lower relative to the standard public goods game with punishment. There is less cooperation and greater free-riding in the undemocratic punishment environment, especially when the undemocratic environment follows after subjects have already experienced the standard public goods game with democratic punishment at the outset. There is much greater alleviation offered to punished subjects when the undemocratic punishment environment comes first followed by the democratic environment. The degree of alleviation in the undemocratic punishment environment is significantly lower when subjects have already had experience with the democratic punishment environment prior to this.

2.3. Papers in Section 3

This section consists of four papers that I felt applied the insights from social dilemma research to a series of interesting questions. In the first paper of this section, Davide Dragone, Fabio Galeotti and Raimondello Orsini conduct an artefactual field experiment to compare the individual preferences and propensity to contribute to a public good of three pools of subjects: undergraduate students, temporary workers and permanent workers. As far as the workers are concerned, our *ex ante* assumption would be that permanent workers, who have greater stability of employment, would exhibit higher cooperation levels compared to temporary workers, whose relations are short-term and marked by greater turn-over. However, the authors find that

students are more selfish and contribute less than workers, whether temporary or permanent. Temporary and permanent workers have similar other-regarding preferences and display analogous contribution patterns to the public good. The authors suggest that the different contractual arrangements, and the consequent economic and psychological effects, play a minor role compared to other factors. It is likely that socialization and learning on the job (something that is not available to the students) create feelings of community that may have induced both temporary and permanent workers to behave similarly and avoid free-riding opportunities. Also, in-group and mutual insurance considerations may have driven both types of workers to contribute to the public good.

The next three papers do not look at social dilemma games in the strictest sense of the term. However, all of them look at problems that have features of a social dilemma and as such constitute interesting applications of the social dilemma paradigm. In the second paper of this section, Ernan Haruvy and Peter T. L. Popkowski Leszczyc look at behavior in auctions with a price externality. These are auctions where only one bidder wins the auction but all bidders derive utility from the winning price. Such include charity auctions where the benefit to the bidders is relatively direct such as auctions to raise money for schools or churches; with the proceeds going towards enhancing services to members who all bid at the auction regardless of who eventually wins it.

One of the variables that the investigation focused on is the *multiplier*, which represents the degree to which the proceeds from the auction directly affect the bidders. The multiplier here is analogous to the implied marginal per capita returns from the public good as in prior studies. The authors report that bidders bid significantly below their valuations in most conditions and well below the theoretical optimal bid. The empirical findings in the laboratory stand in sharp contrast to theoretical predictions, which suggest that bidders should bid more than their valuations and that these bids should increase in the multiplier. The experimental data from this study show that the bidders facing a moderate to high multiplier do not bid aggressively, which can be explained by bidders perceiving higher benefit to losing than to winning. However, increasing the number of bidders reduces the extent of under-bidding. Moreover, despite a substantial level of underbidding, as the multiplier increases winning prices significantly increase. This is an encouraging finding since it suggests that revenues can be increased, relative to no-externality settings, by conducting auctions with price externality and charitable bidders. Clearly these findings have interesting and implementable insights into the designing of charity auctions.

In the next paper of this section, Israel Waichman, Ch'ng Kean Siang, Till Requate, Aric Shafran, Eva Camacho-Cuena, Yoshio Iida and Shosh Shahrabani undertake a cross-country study of reciprocity in labor markets using the well-known

bilateral gift exchange paradigm, which is designed to simulate labor market interaction between employers and employees. This line of work was introduced in a series of papers written by Ernst Fehr and co-authors including Fehr, Gächter, and Kirchsteiger (1997) [41], Fehr, Kirchler, Weichbold, and Gächter (1998) [42], Fehr, Kirchsteiger, and Riedl, (1993, 1996, 1998) [43–45]. Waichman *et al.* extend this line of work to compare the behavior of subjects from five high-income OECD countries: Germany, Spain, Israel, Japan and the USA. This, in itself, adds value since till date no one has undertaken such an ambitious cross-country study of the existence or lack thereof of labor market reciprocity.

Besides exploring whether there are systematic differences in behavior, the authors further examine if any differences arising can be explained by appealing to differing cultural norms in those countries. They report that in all countries, effort levels are increasing while rejection rates are decreasing in wage offers. The authors document some differences arising purely from matching method: fixed matching versus random re-matching. However, there are some stable differences in behavior across countries in both one-shot and repeated relationships, with the most striking differences being those between Germany and Spain. Germans offer the highest wages, while offers are lowest in Spain. The efficiency wage hypothesis—that higher than equilibrium wage offers are reciprocated by greater than minimum effort levels—is confirmed in all countries expect for Spain, where at best it is only weakly supported by the data. On average, overall surplus generated is highest for German subjects and lowest for Spanish subjects. Finally, German subjects also perform better than those of the other countries in terms of effort and overall surplus under fixed matching.

In addition, the authors use the data generated in each country to estimate the parameter measuring aversion to advantageous inequity following the inequity aversion model of Fehr and Schmidt (1999) [46]. While the authors do not observe large differences in aversion to advantageous inequity in Germany, Israel, Japan and the USA, this aversion is considerably lower in Spain. Waichman *et al.* suggest that prevailing labor market norms may at least partially explain these differences. For instance, the fact that Germans offer higher wages and effort than their American counterparts is consistent with the gift exchange game study by Hannan *et al.* (2002) [47], who speculate that country-specific norms explain the lower wages and effort levels of their American subjects in comparison with similar gift-exchange studies conducted by Fehr and colleagues in Europe. Bornhorst *et al.* (2010) [48] also report similar differences between subjects from northern and southern Europe, which is in line with the differences in behavior between German (northern European) and Spanish (southern European) subjects in the current study. These results suggest that the validity of the gift-exchange model and the efficiency wage hypothesis may be culture specific and possibly warrants further examination.

In the final paper of this section Priyodarshi Banerjee, Sujoy Chakravarty and Sanmitra Ghosh also study the impact of one-way, unenforceable pre-play communication in the form of promises like Deer and Bayer (in this volume) except Banerjee *et al.* look at the efficacy of such promises in the context of dictator and ultimatum games. There are two experimental treatments. In the first "no competitive selection" treatment, the proposer in the dictator/ultimatum game makes a promise to the responder about the former's potential offer to the latter. After hearing the promise regarding how much the proposer will offer, the responder can choose not to proceed with the game at all resulting in both players receiving nothing. In a second and more interesting "competitive selection" treatment there are two proposers who each make a promise about their offers to a single responder. After hearing the promise, the responder can choose one of the two proposers to play with; of course, the responder can refuse to play with either. In a control treatment, the game starts with the proposer making a promise to the recipient, regarding the offer. However, here the rest of the game proceeds exactly as in traditional dictator/ultimatum games with the promise playing no role whatsoever; and the responder does not have any power to terminate the game by choosing not to play any further.

The subgame perfect outcome is the same across each treatment of the ultimatum game; the proposer should offer only a small amount to the responder, which the latter should accept and promises should make no difference whatsoever. Furthermore, on the basis of prior research (see, for instance, Hoffman, McCabe, Shachat and Smith, 1994 [49]) we might expect actual offers to be lower with competitive selection because here the proposer who is selected has acquired the "property right" to make an offer. The authors report that, compared to the control treatment, offers are higher in the ultimatum games without or with competitive selection of the proposer. In fact, when selection is competitive, with two proposers at the outset, offers increase even further as opposed to the treatment where selection is not competitive. No such striking differences emerge in the dictator environment, which does not provide an opportunity for the responder to reject offers; thus selection power carries no benefits in the dictator game. Finally, independent of the game institution or proposer selection mechanism, promises provide credible signals of eventual offers, with a strong positive correlation between promises and offers, even though the promise made is merely cheap talk. Offers are highly likely to be rejected, if and when the actual offer falls short of the promise made prior to making that offer. This reinforces prior findings that cheap talk messages can improve cooperation in bargaining games with the additional finding that competitive selection of the proposer emphasizes the salutary effects.

3. Concluding Remarks

The articles in this volume collectively represent the latest advances in how people think of social dilemma problems, how we may be able to enhance cooperation and reduce free-riding in such problems and how we can extend the lessons learned to a host of other similar issues facing us. We have learned, for instance, that a "take" frame does not necessarily lead to lower cooperation compared to a "give" frame but combining a "take" frame with fine-grained individual level feedback leads to more extreme behavior in terms of both greater cooperation and greater free-riding. We have also learned that a strategy based on payoff sampling may provide a more parsimonious and less parameter dependent way of modelling behavior in common pool resource extraction games. We find that people behave differently in social dilemmas when making decisions of their own as opposed to deciding on behalf of someone else.

In some cases the insights, while not radically new, nevertheless provide reassuring corroboration of firmly held priors. For instance, it comes as no surprise that punitive mechanisms lead to reduced free-riding but it is worth learning that such punitive mechanisms fare much better when arrived at via democratic means than autocratic ones. This is true in the case of standard public goods games but is also true in curbing undesirable social practices like cyber-loafing. Similar corroboration is provided for the intuition that cheap-talk messages can be very effective in enhancing cooperation; but equally there may be mediating factors at play such as the distribution of costs between participants in the social dilemma. Similarly, one would assume that cheap-talk messages would be much less effective in a setting with multiple proposers in an ultimatum game. But it turns out that this is not true; even here cheap-talk promises matter and responders pay careful attention to that promise in deciding whether or not to accept the offer.

Finally, we have innovative applications of the social dilemma paradigm to interesting applications such as charity auctions. We have also seen an ambitious study that identifies interesting and interpretable differences across different subject pools and indeed across countries and cultures when it comes to the gift exchange and efficiency wage model. Collectively, the papers in this volume should provide an excellent reference for researchers working in this area, not only in getting a feel for where current thinking stands but also in terms of identifying avenues for future research.

Last, but not the least, I want to extend my sincere gratitude to all the referees who provided reports, often on multiple revisions and always in an expeditious manner. It would have been impossible to put this volume together, in the time-frame that we did, without the generosity and conscientiousness of the referees. I have refrained from thanking them by name in order to preserve confidentiality. I am also

deeply indebted to Haiqing Chen at the *Games* editorial office for her prompt and efficient handling of submissions. She made a demanding job much easier.

References

1. Bohm, P. Revealing demand for an actual public good. *J. Public Econ.* **1983**, *24*, 135–151.
2. Bohm, P. Estimating demand for public goods: An experiment. *Eur. Econ. Rev.* **1972**, *3*, 111–130.
3. Marwell, G.; Ames, R. Economists free ride, does anyone else? *J. Public Econ.* **1981**, *15*, 295–310.
4. Marwell, G.; Ames, R. Experiments on provision of public goods II: Provision point, stake, experience, and the free riding problem. *Am. J. Sociol.* **1980**, *85*, 926–937.
5. Marwell, G.; Ames, R. Experiments on provision of public goods I: Resources, interest, group size, and the free riding problem. *Am. J. Sociol.* **1979**, *84*, 1335–1360.
6. Andreoni, J. Cooperation in public goods experiments: kindness or confusion? *Am. Econ. Rev.* **1995**, *85*, 891–904.
7. Andreoni, J. Impure altruism and donations to public goods: A theory of warm-glow giving. *Econ. J.* **1990**, *100*, 464–477.
8. Andreoni, J. Why free ride? Strategies and learning in public goods experiments. *J. Public Econ.* **1988**, *37*, 291–304.
9. Isaac, M.; Walker, J. Communication and free riding behavior: The voluntary contributions mechanism. *Econ. Inq.* **1988**, *26*, 585–608.
10. Isaac, M.; Walker, J. Group size effects in public goods provision: The voluntary contributions mechanism. *Quart. J. Econ.* **1988**, *103*, 179–199.
11. Isaac, M.; McCue, K.; Plott, C. Public goods provision in an experimental environment. *J. Public Econ.* **1985**, *26*, 51–74.
12. Dawes, R.; McTavish, J.; Shaklee, H. Behavior, communication and assumptions about other people's behavior in a common dilemma situation. *J. Pers. Soc. Psychol.* **1977**, *35*, 1–11.
13. Dawes, R. Social dilemmas. *Annu. Rev. Psychol.* **1980**, *31*, 169–193.
14. Dawes, R.; Orbell, J.; Simmons, R.; van de Kragt, A. Organizing groups for collective action. *Am. Polit. Sci. Rev.* **1986**, *8*, 1171–1185.
15. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*; Cambridge University Press: Cambridge, UK, 1990.
16. Ostrom, E.; Gardner, R.; Walker, J. *Rules, Games, and Common Pool Resources*; University of Michigan Press: Ann Arbor, MI, USA, 1994.
17. Ostrom, E.; Walker, J.; Gardner, R. Covenants with and without a sword: Self-governance is possible. *Am. Polit. Sci. Rev.* **1992**, *86*, 404–417.
18. Yamagishi, T. The provision of a sanctioning system in the United States and Japan. *Soc. Psychol. Quart.* **1988**, *51*, 265–271.
19. Yamagishi, T. The provision of a sanctioning system as a public good. *J. Pers. Social Psychol.* **1986**, *51*, 110–116.

20. Fehr, E.; Gächter, S. Altruistic punishment in humans. *Nature* **2002**, *415*, 137–140.
21. Fehr, E.; Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **2000**, *90*, 980–994.
22. Gintis, H.; Bowles, S.; Boyd, R.; Fehr, E. *Moral Sentiments and Material Interests*; MIT Press: Cambridge, MA, USA, 2005.
23. Hamilton, W. The Genetical Evolution of Social Behavior. *J. Theor. Biol.* **1964**, *37*, 1–52.
24. Trivers, R. The Evolution of Reciprocal Altruism. *Quart. Rev. Biol.* **1971**, *46*, 36–57.
25. Maynard Smith, J. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.
26. Axelrod, R. An evolutionary approach to norms. *Am. Polit. Sci. Rev.* **1986**, *80*, 1095–1111.
27. Palfrey, T.; Prisbrey, J. Anomalous behavior in public goods experiments: How much and why? *Am. Econ. Rev.* **1997**, *87*, 829–846.
28. Anderson, S.; Goeree, J.; Holt, C. A theoretical analysis of altruism and decision error in public goods games. *J. Public Econ.* **1998**, *70*, 297–323.
29. Fischbacher, U.; Gächter, S.; Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **2001**, *71*, 397–404.
30. Gunnthorsdottir, A.; Houser, D.; McCabe, K. Disposition, history and contributions in public goods experiments. *J. Econ. Behav. Organ.* **2007**, *62*, 304–315.
31. Ambrus, A.; Pathak, P. Cooperation over finite horizons: A theory and experiments. *J. Public Econ.* **2011**, *95*, 500–512.
32. Chaudhuri, A.; Paichayontvijit, T.; Smith, A. *Belief Heterogeneity among Conditional Cooperators and the Decay in Contributions in Linear Public Goods Games: Experimental Evidence*; Working Paper; University of Auckland: Auckland, New Zealand, 2015.
33. Fischbacher, U.; Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *Am. Econ. Rev.* **2010**, *100*, 541–556.
34. Neugebauer, T.; Perote, J.; Schmidt, U.; Loos, M. Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *J. Econ. Psychol.* **2009**, *30*, 52–60.
35. Smith, A. Estimating the causal effect of beliefs on contributions in repeated public good games. *Exp. Econ.* **2013**, *16*, 414–425.
36. Ledyard, J. Public Goods: Some Experimental Results. In *Handbook of Experimental Economics*; Kagel, J., Roth, A., Eds.; Princeton University Press: Princeton, NJ, USA, 1995.
37. Chaudhuri, A. Sustaining cooperation in laboratory public goods games: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83.
38. Cárdenas, J.C. Norms from outside and from inside: An experimental analysis on the governance of local ecosystems. *For. Policy Econ.* **2004**, *6*, 229–241.
39. Osborne, M.; Rubinstein, A. Games with Procedurally Rational Players. *Am. Econ. Rev.* **1998**, *88*, 834–847.
40. McKelvey, R.; Palfrey, T. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38.
41. Fehr, E.; Gächter, S.; Kirchsteiger, G. Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica* **1997**, *65*, 833–860.

42. Fehr, E.; Kirchler, E.; Weichbold, A.; Gächter, S. When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets. *J. Labor Econ.* **1998**, *16*, 324–351.
43. Fehr, E.; Kirchsteiger, G.; Riedl, A. Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quart. J. Econ.* **1993**, *108*, 437–459.
44. Fehr, E.; Kirchsteiger, G.; Riedl, A. Involuntary Unemployment and Non-Compensating Wage Differentials in an Experimental Labour Market. *Econ. J.* **1996**, *106*, 106–121.
45. Fehr, E.; Kirchsteiger, G.; Riedl, A. Gift Exchange and Reciprocity in Competitive Experimental Markets. *Europ. Econ. Rev.* **1998**, *42*, 1–34.
46. Fehr, E.; Schmidt, K. A theory of fairness, competition and cooperation. *Quart. J. Econ.* **1999**, *114*, 817–868.
47. Hannan, R.; Kagel, J.; Moser, D. Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior. *J. Labor Econ.* **2002**, *20*, 923–951.
48. Bornhorst, F.; Ichino, A.; Kirchkamp, O.; Schlag, K.; Winter, E. Similarities and differences when building trust: the role of cultures. *Exp. Econ.* **2010**, *13*, 260–283.
49. Hoffman, E.; McCabe, K.; Shachat, K.; Smith, V. Preferences, Property Right and Anonymity in Bargaining Games. *Games Econ. Behav.* **1994**, *7*, 346–380.

Section 1

Framing and Feedback in Social Dilemmas with Partners and Strangers

Caleb A. Cox and Brock Stoddard

Abstract: We study framing effects in repeated social dilemmas by comparing payoff-equivalent Give- and Take-framed public goods games under varying matching mechanisms (Partners or Strangers) and levels of feedback (Aggregate or Individual). In the Give-framed game, players contribute to a public good, while in the Take-framed game, players take from an existing public good. The results show Take framing and Individual-level feedback lead to more extreme behavior (free-riding and full cooperation), especially for Partners. These results suggest Take framing and Individual-level feedback increase the variability of cooperation.

Reprinted from *Games*. Cite as: Cox, C.A.; Stoddard, B. Framing and Feedback in Social Dilemmas with Partners and Strangers. *Games* **2015**, *6*, 394–412.

1. Introduction

There is extensive evidence that behavior in social dilemmas is affected by framing, that is, whether the game is posed in the form of contributing from private accounts to provide a public good ("Give" frame) or extracting from an existing common pool to enrich one's own private account ("Take" frame). However, previous studies show mixed results, with some finding higher cooperation in cases of Give framing, while others find no differences, or even the opposite effect. In this paper, we seek to discover the types of environments in which framing effects are most prevalent, while also helping to explain the mixed results in the literature. Specifically, we examine the nature of the feedback provided, whether subjects are informed about aggregate contribution/extraction ("Aggregate" feedback) or provided disaggregated information at the level of individual choices ("Individual" feedback), and the mechanism used for matching subjects, whether subjects are matched with the same subjects for all rounds ("Partners" matching) or randomly re-matched with different subjects in each round ("Strangers" matching).

Framing effects may differ depending on the level of feedback subjects receive. An individual's action being observed may increase the intensity of Andreoni's [1] "warm glow" or "cold prickle," due to a sense of pride or shame in cooperating or free-riding. Furthermore, framing, feedback, and matching each may affect social norms which arise in the lab, and may interact in interesting ways not yet

well understood. We seek to systematically explore these interactions with our empirical results.[1]

The results of the experiment show Take framing and Individual feedback lead to more extreme behavior (free-riding and full cooperation), especially for Partners. The effect is particularly noticeable for the Partners-Take-Individual (PTI) treatment, where a majority of cooperation decisions are at the extremes. These results suggest Take framing and Individual feedback increase the variability of cooperation. Such high variability could be responsible for the mixed results of previous studies on framing and feedback. We also find gender differences in response to framing, with men responding more negatively to Take framing than women. Such gender differences might also help explain the mixed results in previous studies.

Table 1 summarizes a number of studies of Give/Take framing effects in social dilemmas, while Table 2 summarizes several experiments on feedback.[2] The effect of framing and the effect of feedback on average cooperation are mixed. These previous studies on framing and feedback differ from one another on a number of design features which may partly explain the inconsistent results. To our knowledge, our paper is the first to systematically study the interaction between framing, matching, and feedback conditions to investigate the types of environments in which Give/Take framing effects are most prevalent. The inconsistent previous results may be partly due to differences in such conditions. Moreover, we examine matching and feedback with Take framing, while the previous literature examining cooperation in these environments has primarily focused on Give framing.

The paper is organized as follows. Section 2 discusses the experimental design and hypotheses. Section 3 reports the results. Section 4 concludes with a discussion of the key findings. Experimental instructions and additional analyses are shown in the Appendices.

---

[1] Similar to Andreoni's warm glow and cold prickle explanation is the idea that free-riding may have different "moral costs" depending on framing, level of feedback, and repetition. See Banerjee (2015a [2], b [3]) for further discussion of moral costs.

[2] We focus on studies using a linear VCM game, as in our experiment. See Andreoni and Croson [4] for a summary of previous experiments comparing partners and strangers.

Table 1. Summary of Experiments on Give/Take Framing in Linear VCM Game.

Paper	Feedback	Matching
No Framing Effect		
Fleishman [5]	Aggregate	Partners
Sell and Son [6]	Aggregate & None	Partners
Messer *et al.* [7]	Aggregate	Partners
Cubitt *et al.* [8]	Individual	One-Shot
Cox *et al.* [9]	None	One-Shot
Stoddard [10]	None	Strangers
Cox *et al.* [11]	Individual	Partners
Cooperation Higher in Give		
Andreoni [1]	Aggregate	Strangers
Park [12]	Aggregate	Strangers
Brandts and Schwieren [13]	None	Strangers
Fujimoto and Park [14]	Aggregate	One-Shot
Dufwenberg *et al.* [15]	Aggregate	One-Shot
Gächter *et al.* [16]	None	Strangers
Cox [17]	Aggregate	Strangers
Khadjavi and Lange [18]	Aggregate	Partners
Cooperation Higher in Take		
Sell *et al.* [19]	Aggregate	Partners
Fosgaard *et al.* [20]	Individual	One-Shot

Table 2. Summary of Experiments on Individual/Aggregate Feedback in Linear VCM Game.

Paper	Framing	Matching
No Feedback Effect		
Weimann [21]	Give	Partners & Strangers
Croson [22]	Give	Partners
Bigoni and Suetens [23]	Give	Partners
Cooperation Higher in Individual		
Sell and Wilson [24]	Give	Partners
Kreitmair [25]	Give	Partners
Cooperation Higher in Aggregate		
Van der Heijden and Moxnes [26]	Take	Partners
Carpenter [27]	Give	Strangers

2. Experimental Design and Procedures

The experimental sessions were conducted at the University of South Dakota. Undergraduate subjects from a wide range of disciplines were recruited from classrooms. At the beginning of each session, subjects read a set of instructions, which were then summarized publicly. After reading the instructions, subjects took a post-instruction quiz and were not allowed to continue until all answers were correct.

Subjects made all decisions on computers in private.[3] At the beginning of each session, the computer randomly and anonymously assigned subjects to three-person groups. No person could identify his/her group members. Each session consisted of 15 decision periods.

The first treatment variable is framing (Give *vs.* Take). In the Give frame, each group member received a endowment of 80 tokens each round to allocate between his/her private account and a group account. Each token allocated to the group account produced 2.25 tokens for the group. Each group member received an equal share of the ending value of the group account, *i.e.*, a marginal per-capita return of 0.75 tokens for each token provided to the group account. In the Take frame, by contrast, the group account was endowed with 540 tokens. Each group member could appropriate a maximum of 80 tokens from the group account to his/her private account. Each token appropriated from the group account decreased the group account by 2.25 tokens. Each person received an equal share of the ending value of the group account, *i.e.*, one third of each token remaining in the group account.

The second treatment variable is feedback level (Aggregate *vs.* Individual). Under Aggregate feedback, each subject was informed of his/her group's aggregate contribution to or appropriation from the group account and his/her earnings at the end of each period. Under Individual feedback, subjects were also informed of the other two anonymous group members' individual allocation decisions in each period.

The third treatment variable is matching (Partners *vs.* Strangers). Under Partners matching, subjects were randomly and anonymously assigned into groups of three before the first decision period and remained in the same group for all 15 periods. Under Strangers matching, subjects were randomly and anonymously assigned into new groups of three before each of the 15 decision periods.

Three treatment variables with two levels each gives us a $2 \times 2 \times 2$ design. Treatment conditions are identified by the matching mechanism-framing-feedback level. For instance, the treatment condition with Strangers matching, Give framing, and Aggregate feedback is identified as the Strangers-Give-Aggregate treatment (or SGA). Table 3 lists each treatment and summarizes its design. Most sessions had 15 subjects, though one session had only 12 due to absences. In total, data were collected from 252 subjects.[4] In all sessions, monetary information was denominated in tokens.

---

[3] Co-author Cox programmed the experiment using z-Tree Fischbacher [28]. Co-author Stoddard conducted all sessions, including a review of the instructions. See the Appendices for a copy of the instructions. Note, following the previous economics literature on framing in social dilemmas, steps were taken to minimize unwanted demand effects. We used a between-subjects design, and instructions avoided terms such as "give" and "take" in favor of more neutral language.

[4] As we will show in the Results section, the data from the first two sessions in each treatment suggested more complete free-riding and more full cooperation in PTI treatment. We had funds remaining for

The conversion rate of tokens to U.S. dollars was 100 to 1. Earnings averaged $18.74 per subject across all periods, which ranged in duration from 45–60 min.

Table 3. Design information for treatments.

Treatment	Sessions	Independent Groups	Subjects (Men/Women)
Partners-Give-Aggregate (PGA)	2	10	30 (15/15)
Partners-Give-Individual (PGI)	2	10	30 (13/17)
Partners-Take-Aggregate (PTA)	2	10	30 (14/16)
Partners-Take-Individual (PTI)	3	15	45 (32/13)
Strangers-Give-Aggregate (SGA)	2	2	30 (9/21)
Strangers-Give-Individual (SGI)	2	2	27 (20/7)
Strangers-Take-Aggregate (STA)	2	2	30 (8/22)
Strangers-Take-Individual (STI)	2	2	30 (12/18)

Predictions

On the basis of prior research findings, we state the following hypotheses. As summarized in Table 1, related studies find mixed results regarding the effect of framing on average cooperation. However, among papers that find a significant framing effect, cooperation tends to be higher with Give framing.

Hypothesis 1: Between treatments with the same matching mechanism and feedback condition, average cooperation will be higher with Give framing compared to Take framing.

Furthermore, the framing effect may be sensitive to gender. Fujimoto and Park [14] and Cox [17] find higher cooperation in Give compared to Take, with a stronger effect for men than women.[5]

Hypothesis 2: Between treatments with the same matching mechanism and feedback condition, the framing effect will be stronger for men than for women.

The literature studying information feedback in public-goods games also finds mixed results on average cooperation, as summarized in Table 2. However, cooperative decisions have more variance with Individual feedback (Croson [22], Bigoni and Suetens [23]).

Hypothesis 3: Between treatments with the same matching mechanism and frame condition, there will be more extreme behavior (free riding and full cooperation) with Individual feedback compared to Aggregate feedback.

---

one additional session, and we chose PTI to see if these results replicated. Our results show that PTI continues to exhibit a greater degree of such extreme behavior.

[5] One possible reason why men may be more responsive to Take framing is it may prime competitive norms and men often have a stronger response to competition than women (Croson and Gneezy [29]).

Studies examining matching mechanisms with Give framing do not find consistent results, and to our knowledge, the effect of matching mechanisms has not been examined with Take framing. It is thus unclear ex ante what effect of matching should be expected.

3. Results

The results from the experiment are discussed in two subsections according to the matching mechanism.[6] Within each subsection, the cooperation levels, proportion of free riders, and the proportion of full cooperators are examined. Each analysis begins with a graphical overview, followed with individual-level regressions.[7] Take and Give decisions are standardized across frames as cooperation decisions. A cooperation decision in Take framing equals 80 tokens minus the tokens appropriated from the group account. A cooperation decision in Give framing equals the number of tokens provided to the group account.

Recall the treatments are identified by matching mechanism-frame-feedback condition. For brevity in the discussion of the results, the Partners-Give-Aggregate treatment will be referred to as PGA, the Partners-Give-Individual treatment as PGI, the Partners-Take-Aggregate treatment as PTA, and the Partners-Take-Individual treatment as PTI. Similar acronyms are used for the Strangers treatments.

3.1. Partners-Matching Data

3.1.1. Graphical Overview: Partners

Figure 1 displays the path of average individual-level cooperation for each treatment. The pattern of average cooperation across decision periods is similar in all four treatments. Holding constant the feedback condition, the Take treatments have higher levels of average cooperation. Similarly, holding constant the frame, the Individual-feedback treatments have higher levels of average cooperation. Focusing on the first period, visually it appears there is a framing effect where Take framing leads to higher average cooperation than Give framing. This observation is supported by two-sample t-tests (PGA *vs.* PTA $p = 0.030$, $n = 60$; PGI *vs.* PTI $p = 0.066$, $n = 75$).[8]

---

[6] We also compare cooperative behavior between Partners and Strangers matching for given frame and feedback conditions. The results do not find strong differences between matching conditions, and are reported in the Appendices. To reduce the size of the Results section, we also report descriptive statistics, proportions of free riders, and proportions of full cooperators in the Appendices.

[7] In regression tables, we report statistical significance for 2-tailed tests. However, as our hypotheses are 1-sided, we will also report 1-tailed tests in the text where relevant.

[8] Wilcoxon Rank-sum tests support the significant difference between PGA *vs.* PTA, but not a significant difference between PGI *vs.* PTI ($p = 0.180$). The insignificant Wilcoxon test is likely due to the higher proportions of extreme behavior observed with Individual feedback. The full set of treatment comparisons and tests are included in the Appendices.

Result 1: Contrary to Hypothesis 1, first period average cooperation is higher with Take framing compared to Give framing.

Previous research indicates average cooperation can mask a large degree of variation in cooperation decisions. Some of the variation is caused by extreme behavior of subjects, which has been shown to be sensitive to Take framing (Cox [17]) and Individual feedback (Croson [22], Kreitmair [25]). Figure 2 displays the path of the proportion of complete free riders for each treatment. Within a feedback condition, there is always a higher proportion of free riders with Take framing.

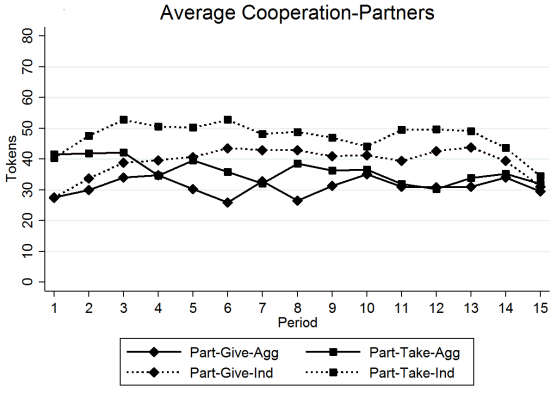


Figure 1. Average Cooperation Over Time (Partners).

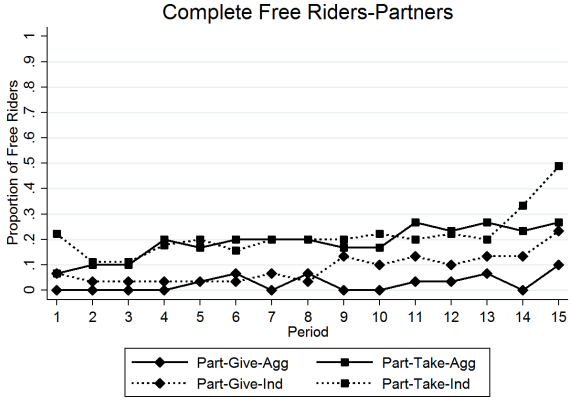


Figure 2. Proportion of Free Riders Over Time (Partners).

Finally, a group's total payoff is maximized when each individual member fully cooperates (coop = 80). Figure 3 displays the path of the proportion of full cooperators for each treatment. There are not large differences in the proportions of full cooperation between treatments, except for the much higher proportion of full cooperators in the PTI treatment.

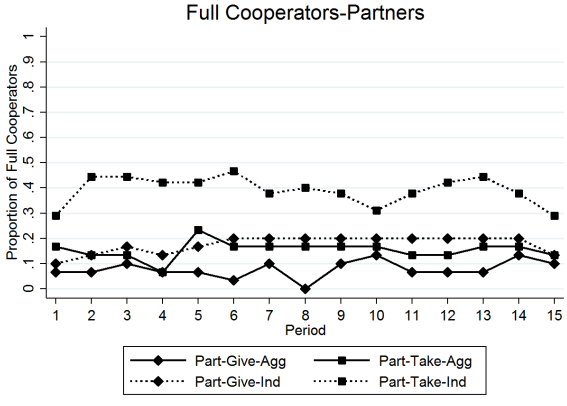


Figure 3. Proportion of Full Cooperators Over Time (Partners).

3.1.2. Individual-Level Cooperation Regression Analysis: Partners

Models 1 and 2 in Table 4 present results from two-limit censored Tobit regression models. The dependent variable in each model is individual-level cooperation. Each model pools data from the four Partners-matching treatments across all 15 periods. The independent variables in Model 1 are treatment indicators (with PGA as the reference category), a female gender indicator, and a time-trend variable. Model 2 has the same independent variables as Model 1, as well as an additional independent variable, the one-period lagged average cooperation of other group members.[9]

Model 1 reveals a strong positive effect on cooperation with Take framing and Individual feedback, PTI coefficient (p-value = 0.031). Consistent with other previous linear public good studies, Model 2 reports lagged average cooperation of others has a significant and positive impact on individual cooperation decisions, (p-value < 0.001). Including lagged cooperation of others decreases the size and significance of the effect of Take framing and Individual Feedback, PTI coefficient (p-value = 0.142 in Model 2). In addition, common to public goods studies, Model 2

[9] Lagged cooperation decisions of others influence contributions to public goods and control for feedback effects (Sefton *et al.* [30], Frechette [31], Samek and Sheremeta [32]).

reports a significant and negative time trend of cooperation, Period (p-value < 0.001 in Model 2). There is not a gender effect with average cooperation decisions in either model (p-values = 0.642 in Model 1 and 0.488 in Model 2). Finally, more than a third of the observations in each model are censored due to free-riding or full cooperation.[10]

Table 4. Individual-Level Regressions: Average Cooperation, Free-Riding, and Fully Cooperating (Partners).

Independent Variables	Cooperation		Free-Riding		Fully Cooperating	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Coeff.	Coeff.	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
Constant	36.71 ***	15.10 ***	0.02 ***	0.03 ***	0.11 ***	0.04 ***
	(5.44)	(4.57)	(0.01)	(0.01)	(0.05)	(0.02)
Lagged Average Cooperation of Others	—-	0.79 ***	—-	0.98 **	—-	1.04 ***
	—-	(0.14)	—-	(0.01)	—-	(0.01)
PTA	–2.98	–1.51	9.50 ***	10.06 ***	2.23	1.84
	(9.44)	(4.45)	(3.49)	(3.12)	(1.54)	(1.02)
PGI	9.48	3.93	3.82 **	4.11 ***	2.70	2.00
	(10.36)	(4.56)	(2.15)	(2.02)	(1.78)	(1.10)
PTI	20.65 **	7.84	8.88 ***	11.63 ***	6.77 ***	4.21 ***
	(9.58)	(5.33)	(3.76)	(4.41)	(3.45)	(1.86)
Female	–2.09	–2.60	0.34 ***	0.34 ***	0.46 ***	0.40 ***
	(4.50)	(3.74)	(0.13)	(0.13)	(0.14)	(0.13)
Period	–0.45	–0.73 ***	1.11 ***	1.11 ***	1.00	0.99
	(0.38)	(0.23)	(0.02)	(0.02)	(0.02)	(0.02)
Pseudo R^2	—-	—-	0.127	0.171	0.106	0.375
Wald Tests	p-value	p-value	p-value	p-value	p-value	p-value
PTA=PTI	0.126	0.178	0.863	0.657	0.092*	0.135
PGI=PTI	0.366	0.542	0.142	0.027**	0.143	0.156

Standard errors are in parentheses. In all models, errors are clustered by group (45). ***, **, and * indicate significance at the 1%, 5%, and 10% levels using 2-tailed tests. Partners-Give-Aggregate (PGA) serves as the reference treatment indicator variable in all models; Model 1: n = 2025, 282 left-censored (coop = 0) & 447 right-censored (coop = 80) observations; Model 2: n = 1890; 268 left-censored (coop = 0) & 424 right-censored (coop = 80) observations; Model 3: n = 2025; Model 4: n = 1890; Model 5: n = 2025; Model 6: n = 1890.

[10] As a robustness check for Models 1 and 2, models with the same dependent and independent variables are examined using pooled OLS with Driscoll and Kraay (DK) standard errors, rather than clustered standard errors. Driscoll and Kraay [33] developed standard errors that are robust to arbitrary spatial and serial correlation. Vogelsang [34] shows that for panels with large time dimensions and finite cross-sectional dimensions, the DK standard errors are consistent even in specifications with time fixed effects. The estimation procedure developed by Hoechle [35] is implemented in STATA. The results from the specifications with DK standard errors are generally consistent with those from the Tobit models reported in Table 4.

Result 2: Contrary to Hypothesis 1, average cooperation is not higher with Give framing compared to Take framing, and there is some evidence the opposite is true with Individual feedback.

3.1.3. Free-Riding Regression Analysis: Partners

Models 3 and 4 in Table 4 report logit regressions for complete free-riding (coop = 0).[11] The dependent variable in each model is an indicator variable equal to one if a subject's cooperation decision in a particular period is 0. The independent variables in Models 3 and 4 are the same as those in Models 1 and 2, respectively. There is significantly higher likelihood of free-riding with Take framing (p-values < 0.001 in Model 3 and Model 4). With Give framing, Individual feedback leads to significantly higher likelihood of free-riding (1-tailed p-value < 0.001 in Model 3 and Model 4). With Take framing, however, Wald tests comparing PTA and PTI indicate Individual feedback does not lead to significantly higher likelihood of free-riding (1-tailed p-values = 0.432 in Model 3 and 0.329 in Model 4). Women are significantly less likely to free ride compared to men (p-value < 0.001 in Model 3 and Model 4). Finally, the time-trend odds ratios are significant and greater than one (p-values < 0.001 in Model 3 and Model 4), indicating free-riding is more likely to occur in later periods of the game.

3.1.4. Full Cooperation Regression Analysis: Partners

Models 5 and 6 in Table 4 report logit regressions for full cooperation (coop = 80). The dependent variable in each model is an indicator variable equal to one if a subject's cooperation decision in a particular period is 80. The independent variables in Models 5 and 6 are the same as those in Models 1 and 2, respectively. Full cooperation is more likely in PTI combining Take framing with Individual feedback compared to the other treatments (p-value < 0.001 in Model 5 and Model 6). Wald tests comparing PTA and PTI indicate full cooperation with Take framing is weakly more likely with Individual feedback compared to Aggregate feedback (1-tailed p-values = 0.046 in Model 5 and 0.068 in Model 6). Women are much less likely to fully cooperate than men (p-value < 0.001 in Model 5 and Model 6).[12] Finally, time trends do not significantly impact the likelihood of subjects' full cooperation (p-values = 0.862 in Model 5 and 0.453 in Model 6).

---

[11] Note that for logit regressions we report odds ratios, so that estimates greater than 1 indicate a positive effect, and estimates less than 1 indicate a negative effect.

[12] The full-cooperation rate for women in the PGA treatment was precisely zero.

Result 3: Supporting Hypothesis 3, with Give framing there is significantly more free-riding with Individual feedback compared to Aggregate feedback. However, such a feedback effect is not present with Take framing. Examining full cooperation, there is weak support of Hypothesis 3 with Take framing, but no support with Give framing. Free-riding and full cooperation increase significantly when Take framing and Individual feedback are combined.

3.2. Strangers-Matching Data

3.2.1. Graphical Overview: Strangers

Figure 4 displays the path of average individual-level cooperation for each treatment. The pattern of average group-level cooperation across decision periods is similar in all four treatments. Focusing on the first period, visually it appears there is a framing effect where Take framing leads to higher average cooperation than Give framing. This observation is supported by two-sample *t*-tests (PGA *vs.* PTA p-value = 0.011, n = 60; PGI *vs.* PTI p-value = 0.022, n = 57).[13]

Result 4: Contrary to Hypothesis 1, first period average cooperation is higher with Take framing compared to Give framing (similar to *Result 1* with Partners matching).

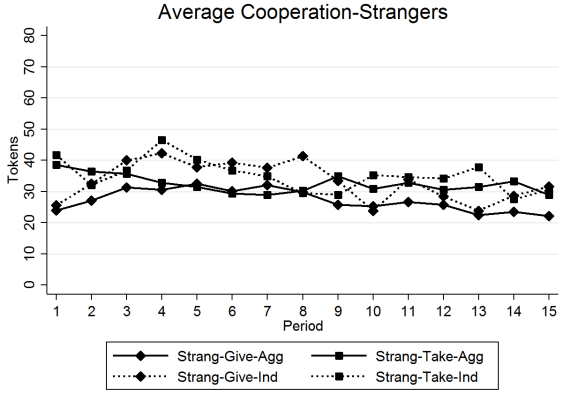


Figure 4. Average Cooperation Over Time (Strangers).

To investigate the variation in cooperative decisions due to extreme behavior, we examine the proportions of complete free riders and full cooperators. Figure 5

[13] Wilcoxon Rank-sum tests also support this observation. The full set of treatment comparisons and tests are included in the Appendices.

displays the path of the proportions of complete free riders for each treatment. The pattern of the proportions of free riders across decision periods are similar across SGA, STA, and SGI. However, for the last ten periods, the proportion of free riders in STI is above the other three treatments.

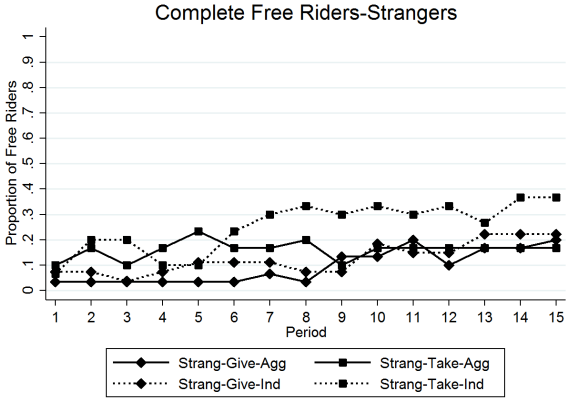


Figure 5. Proportion of Free Riders Over Time (Strangers).

Figure 6 displays the path of the proportions of full cooperators for each treatment. The pattern of the proportions of full cooperators across decision periods are higher in treatments with Individual feedback compared to treatments with Aggregate feedback.

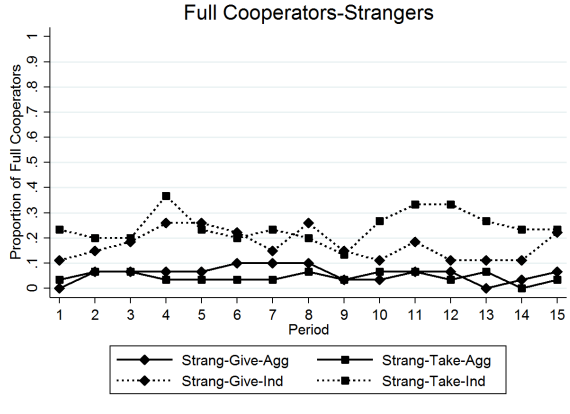


Figure 6. Proportion of Full Cooperators Over Time (Strangers).

3.2.2. Individual-Level Cooperation Regression Analysis: Strangers

Models 1 and 2 in Table 5 present results from two-limit censored Tobit regression models. The dependent variable in each model is individual-level cooperation. Each model pools data from the four Strangers-matching treatments across all 15 periods. Independent variables in Model 1 are treatment indicators (with SGA as the reference category), a female gender indicator, and a time-trend variable. Model 2 has the same independent variables as Model 1, as well as an additional independent variable, the one-period lagged average cooperation of other group members.

Table 5. Individual-Level Regressions: Average Cooperation, Free-Riding, and Fully Cooperating (Strangers).

Independent Variables	**Cooperation**		**Free-Riding**		**Fully Cooperating**	
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Coeff.	**Coeff.**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**
Constant	35.54***	27.22***	0.08***	0.12***	0.12***	0.09***
	(6.50)	(7.18)	(0.04)	(0.06)	(0.08)	(0.06)
Lagged Average	—-	0.28***	—-	0.99**	—-	1.01***
Cooperation of Others	—-	(0.08)	—-	(0.01)	—-	(0.01)
STA	3.92	1.72	1.94	2.03	0.79	0.69
	(6.38)	(6.42)	(1.14)	(1.15)	(0.64)	(0.54)
SGI	5.78	5.20	1.02	1.05	2.17	2.00
	(7.61)	(7.81)	(0.66)	(0.65)	(1.43)	(1.35)
STI	7.86	4.81	3.18**	3.57***	4.96***	4.12***
	(7.13)	(7.34)	(1.71)	(1.85)	(2.72)	(2.33)
Female	–4.32	–3.05	0.47*	0.45*	0.32***	0.36**
	(6.42)	(6.59)	(0.20)	(0.19)	(0.14)	(0.17)
Period	–0.80***	–0.78***	1.10***	1.08***	1.00	0.99
	(0.24)	(0.26)	(0.02)	(0.02)	(0.02)	(0.02)
Pseudo R^2	—-	—-	0.070	0.076	0.121	0.124
Wald Tests	*p*-value	*p*-value	*p*-value	*p*-value	*p*-value	*p*-value
STA = STI	0.624	0.707	0.287	0.220	0.011***	0.012***
SGI = STI	0.814	0.966	0.044**	0.031**	0.110	0.168

Standard errors are in parentheses. In all models, errors are clustered by subject (117). ***, **, and * indicate significance at the 1%, 5%, and 10% levels using 2-tailed tests. Strangers-Give-Aggregate (SGA) serves as the reference treatment indicator variable in all models. Model 1: n = 1755, 279 left-censored (coop = 0) & 226 right-censored (coop = 80) observations; Model 2: n = 1638, 271 left-censored (coop = 0) & 215 right-censored (coop = 80) observations; Model 3: n = 1755; Model 4: n = 1638; Model 5: n = 1755; Model 6: n = 1638.

Models 1 and 2 report no significant treatment effects for cooperation decisions. However, the lagged cooperation of others increases cooperation, Lagged Average Cooperation of Others (p-value < 0.001 in Model 2). Moreover, time trends significantly reduce cooperation (p-value < 0.001 in Model 1 and Model 2). There is not a gender effect with average cooperation decisions in either model

(p-values = 0.501 in Model 1 and 0.644 in Model 2). Finally, greater than a fourth of the observations in each model are censored due to free-riding and full cooperation. [14]

Result 5: Contrary to Hypothesis 1, average cooperation is not higher with Give framing compared to Take framing (similar to *Result 2* with Partners matching).

3.2.3. Free-Riding Regression Analysis: Strangers

Models 3 and 4 in Table 5 report logit regressions for complete free-riding (coop = 0) for Strangers treatments. The dependent variable in each model is an indicator variable equal to one if a subject's cooperation decision in a particular period is 0. Independent variables in Models 3 and 4 are the same as those in Models 1 and 2, respectively. Free-riding is more likely in STI combining Take framing and Individual feedback compared to the other treatments (p-values = 0.031 in Model 3 and p-value = 0.014 in Model 4). With Individual feedback, Wald tests comparing SGI and STI indicate Take framing increases the likelihood of free-riding compared to Give framing (p-values = 0.044 in Model 3 and 0.031 in Model 4). However, with Take framing, Wald tests comparing STA and STI indicate Individual feedback does not lead to significantly higher likelihood of free-riding (1-tailed p-values = 0.144 in Model 3 and 0.110 in Model 4). Women are weakly less likely to free ride compared to men (p-values = 0.077 in Model 3 and 0.056 in Model 4). Finally, time-trend odds ratios are significant and greater than one (p-value < 0.001 in Model 3 and Model 4), indicating free-riding is more likely to occur in later periods of the game.

3.2.4. Full Cooperation Regression Analysis: Strangers

Models 5 and 6 in Table 5 report logit regressions for full cooperation (coop = 80). The dependent variable in each model is an indicator variable equal to one if a subject's cooperation decision in a particular period is 80. The independent variables in Models 5 and 6 are the same as those in Models 1 and 2, respectively. Full cooperation is more likely in STI combining Take framing with Individual feedback compared to the other treatments (p-values = 0.004 in Model 5 and 0.012 in Model 6). Wald tests comparing STA and STI indicate full cooperation with Take framing is more likely with Individual feedback compared to Aggregate feedback (1-tailed p-values < 0.001 in Model 5 and Model 6). Women are less likely to fully cooperate

[14] Strangers matching leads to dependency across subjects within a session. In Table 5, standard errors are clustered by individual subjects, with lagged cooperation of others used to control for feedback effects. Another method for controlling session-level dependency is to cluster by session (Frechette [31]). In cases where there are low numbers of clusters, the DK standard errors provide an attractive alternative to clustered standard errors. Pooled OLS models with DK standard errors and the same dependent and independent variables as Models 1 and 2 are conducted as robustness checks for the results in Table 5. The general results are consistent across both specifications.

compared to men (p-values = 0.011 in Model 5 and 0.027 in Model 6). Finally, time trends do not significantly impact the likelihood of subjects' full cooperation (p-values = 0.814 in Model 5 and 0.652 in Model 6).

Result 6: Contrary to Hypothesis 3, holding constant the framing condition, there is not more free-riding with Individual Feedback compared to Aggregate feedback (unlike *Result 3* with Give framing and Partners matching). Examining full cooperation, there is support of Hypothesis 3 with Take framing, but no support with Give framing (similar to *Result 3* with Partners matching). However, free-riding and full cooperation increase significantly when Take framing and Individual feedback are combined (similar to *Result 3* with Partners matching).

3.3. Framing Effects by Gender

3.3.1. Graphical Overview: Partners

Figure 7 displays the path of average cooperation for men and women by treatment. It appears average cooperation levels for women may vary more by treatment condition than average cooperation levels for men. For instance, women on average cooperate more in PTA with Take framing and Aggregate feedback compared to PGA with Give framing and Aggregate feedback.

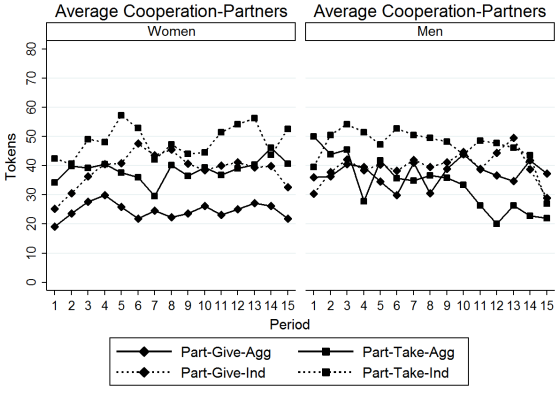


Figure 7. Average Cooperation Over Time (Partners).

3.3.2. Graphical Overview: Strangers

Figure 8 displays the path of average individual-level cooperation for men and women by treatment. For women, the most clear observation is average cooperation with Aggregate feedback is always higher in STA with Take framing than in SGA with Give framing. However, for men, it appears the framing effect with Aggregate

feedback is reversed. That is, in most periods, average cooperation is higher in SGA with Give framing than in STA with Take framing.

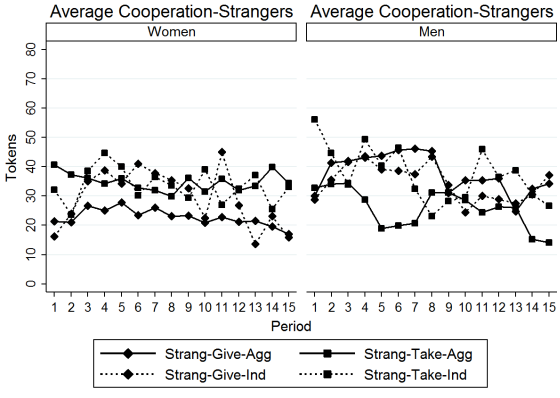


Figure 8. Average Cooperation Over Time (Strangers).

In Table 6 we test Hypothesis 2 by examining the interaction between gender and Take framing. Unlike previous regressions, we use indicators for Take framing, Individual feedback, and Partners matching instead of indicators for individual treatments (such as STA, PGI, *etc.*). In Model 3 in Table 6 we fully interact these treatment variables with one another, which is equivalent to controlling for a full set of treatment indicators (with SGA as the reference category). This approach simplifies the interpretation of the main effect of Take framing and its interaction with gender, compared to interacting gender with treatment condition indicators.

The main effect of Take framing measures the framing effect for men. It is negative, showing only weak statistical significance (1-tailed p-value = 0.090 in Model 3). Women are significantly less cooperative than men. Moreover, the interaction between Take framing and female gender is positive and significant (1-tailed p-values = 0.058 in Model 1, 0.036 in Model 2, and 0.018 in Model 3). This result suggests, consistent with Hypothesis 2, the effect of Take framing differs by gender in the predicted direction. There is some weak evidence that the overall effect of Take framing for women (a linear combination of the Take framing coefficient and the interaction coefficient) is positive (p-values = 0.046 for Model 1, 0.109 for Model 2, and 0.629 for Model 3). While we cannot show strong evidence of a significant Take framing effect for either men or women separately, we do find the effect of Take framing is different for men and women.

Result 7: Take framing weakly reduces cooperation for men and weakly increase cooperation for women. Consistent with Hypothesis 2, Take framing interacts significantly with gender.

Table 6. Tobit Regressions for Cooperation Interacting Gender with Take Framing.

Independent Variable	Model 1	Model 2	Model 3
Constant	33.06 ***	18.51 ***	23.87 ***
	(4.85)	(4.99)	(5.89)
Lagged Cooperation of Others	—-	0.59 ***	0.58 ***
		(0.11)	(0.11)
Take	–0.87	–4.88	–11.59
	(6.75)	(5.37)	(8.63)
Individual	9.70 **	6.94 *	0.55
	(4.80)	(3.85)	(7.78)
Partners	10.06 **	6.69 *	1.59
	(4.72)	(3.77)	(4.35)
Take × Individual	—-	—-	4.30
			(11.51)
Take × Partners	—-	—-	3.40
			(8.63)
Individual × Partners	—-	—-	5.37
			(9.86)
Take × Individual × Partners	—-	—-	3.09
			(14.91)
Female	–9.28 **	–8.68 **	–9.71 **
	(4.37)	(3.88)	(4.08)
Female × Take	11.45	11.96 *	14.57 **
	(7.29)	(6.62)	(6.95)
Period	–0.61 ***	–0.78 ***	–0.78 ***
	(0.23)	(0.20)	(0.20)

Standard errors are in parentheses. In all models, errors are clustered by subject for Strangers and group for Partners (162). ***, **, and * indicate significance at the 1%, 5%, and 10% levels using 2-tailed tests; Model 1: n = 3780, 561 left-censored (coop = 0) & 673 right-censored (coop = 80) observations; Models 2 & 3: n = 3092, 473 left-censored (coop = 0) & 582 right-censored (coop = 80) observations.

4. Discussion

In this paper, we explore how framing, feedback, and matching interact in determining cooperation in social dilemmas. The results indicate that Take framing and Individual feedback lead to more extreme behavior (free-riding and full cooperation). This effect is especially pronounced for Partners matching. In the treatment combining these three features, PTI, more than half of all cooperation decisions are on the extremes.

One possible conjecture to explain these findings is Take framing, Individual feedback, and Partners matching are conducive to the formation of norms for full cooperation, especially in combination. As suggested by Andreoni [1], taking may be viewed more negatively than not giving. This "omission bias" effect may be especially strong when individual cooperation decisions are revealed to fixed partners. In such cases, norm breakers may be inclined to free ride completely, since taking even a small amount still breaks the norm. This result may be explained by the "Broken Window Theory" (Wilson and Kelling [36], Keizer *et al.* [37]). This theory suggests a

broken window in a building, *i.e.*, a signal of another's non-compliance to a social norm, will lead to more broken windows and, if gone unpunished, could escalate to larger crimes in a community. Further investigation of such norm formation and how procedural details affect it is an interesting direction for future research.[15]

Our findings may help explain the mixed results of the previous literature on framing, feedback, and matching in public goods games and related social dilemmas. Extreme behavior at both ends of the strategy space leads to high variance in cooperation. Such high variance could lead to very different results in small samples typical of experiments. Furthermore, our results suggest interactions between these design features, so one design choice (e.g., level of feedback) may affect the observed treatment effect of another variable (e.g., framing). Critically, this finding suggests framing effects may exist, but are not always observed in other studies that focus on average cooperation as the primary variable of interest.

Finally, like Fujimoto and Park [14] and Cox [17], our finding of heterogeneous treatment effects for men and women suggests uncontrolled gender differences might partly explain previous mixed results. Other uncontrolled variables, as well as differences in statistical power, may also be part of the explanation. Nonetheless, our results suggest researchers studying framing, feedback, and matching in social dilemmas should be aware of how these design features may interact, as well as the potential for heterogeneous treatment effects by gender.

Acknowledgments: The authors would like to thank James Walker, Garret Ridinger, Erik Kimbrough, Nick Feltovich, Ursula Kreitmair, and the participants of the 2014 Annual Conference for the Society for the Advancement of Behavioral Economics and the 2014 Economic Science Association European Meeting for helpful comments. We are also very thankful for helpful comments from the editor and anonymous referees. Any remaining errors are the responsibility of the authors. Financial support for this research was generously provided by the Beacom Opportunity Fund at the University of South Dakota.

Author Contributions: Both authors contributed equally to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Andreoni, J. Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. *Q. J. Econ.* **1995**, *110*, 1–21.
2. Banerjee, R. On the interpretation of bribery in a laboratory corruption game: Moral frames and social norms. *Exp. Econ.* **2015**, in press.

[15] Kimbrough and Vostroknutov [38] elicit norms in a public goods game (similar to our PGA treatment), finding evidence of a conditional cooperation norm.

3. Banerjee, R. *Corruption, Norm Violation and Decay in Social Capital*; Economics Working Papers; School of Economics and Management, University of Aarhus: Aarhus, Denmark, 2015.
4. Andreoni, J.; Croson, R. Partners *vs.* Strangers: Random Rematching in Public Goods Experiments. In *Handbook of Experimental Economics Results*; Plott, C.R., Smith, V.L., Eds.; North-Holland: Amsterdam, The Netherlands, 2008; pp. 776–783.
5. Fleishman, J.A. The Effects of Decision Framing and Other's Behavior on Cooperation in a Social Dilemma. *J. Confl. Resolut.* **1988**, *32*, 162–180.
6. Sell, J.; Son, Y. Comparing Public Goods with Common Pool Resources: Three Experiments. *Soc. Sci. Q.* **1997**, *60*, 118–137.
7. Messer, K.; Zarghamee, H.; Kaiser, H.; Schulze, W. New hope for the voluntary contributions mechanism: The effects of context. *J. Public Econ.* **2007**, *91*, 1783–1799.
8. Cubitt, R.P.; Drouvelis, M.; Gächter, S. Framing and free riding: Emotional responses and punishment in social dilemma games. *Exp. Econ.* **2011**, *14*, 254–272.
9. Cox, J.C.; Ostrom, E.; Sadiraj, V.; Walker, J.M. Provision versus Appropriation in Symmetric and Asymmetric Social Dilemmas. *South. Econ. J.* **2013**, *79*, 496–512.
10. Stoddard, B. *Uncertainty in Payoff-Equivalent Appropriation and Provision Games*; Working Paper; University of South Dakota: Vermillion, SD, USA, 2014.
11. Cox, J.C.; Sadiraj, V.; Walker, J.M. *Provision versus Appropriation in Symmetric and Asymmetric Social Dilemmas: Social History and Repetition*; Working Paper; Georgia State University: Atlanta, GA, USA, 2014.
12. Park, E. Warm-glow versus cold-prickle: A further experimental study of framing effects on free-riding. *J. Econ. Behav. Org.* **2000**, *43*, 405–421.
13. Brandts, J.; Schwieren, C. *Frames and Economic Behavior: An Experimental Study*; Working Paper; Universitat Autònoma de Barcelona: Barcelona, Spain, 2009.
14. Fujimoto, H.; Park, E. Framing effects and gender differences in voluntary public goods provision experiments. *J. Socio-Econ.* **2010**, *39*, 455–457.
15. Dufwenberg, M.; Gächter, S.; Hennig-Schmidt, H. The framing of games and the psychology of play. *Games Econ. Behav.* **2011**, *73*, 459–478.
16. Gächter, S.; Kölle, F.; Quercia, S. *The ABC of Cooperation in Voluntary Contribution and Common Pool Extraction Games*; Working Paper; University of Nottingham: Nottingham, UK, 2014.
17. Cox, C. Decomposing the effects of negative framing in linear public goods games. *Econ. Lett.* **2015**, *126*, 63–65.
18. Khadjavi, M.; Lange, A. Doing good or doing harm: Experimental evidence on giving and taking in public good games. *Exp. Econ.* **2015**, *18*, 432–441.
19. Sell, J.; Chen, Z.Y.; Hunter-Holmes, P.; Johansson, A. A cross-cultural comparison of public good and resource good settings. *Soc. Psychol. Q.* **2002**, *65*, 285–297.
20. Fosgaard, T.; Hansen, L.G.; Wengström, E. Understanding the nature of cooperation variability. *J. Public Econ.* **2014**, *120*, 134–143.
21. Weimann, J. Individual behavior in a free riding experiment. *J. Public Econ.* **1994**, *54*, 185–200.

22. Croson, R.T.A. Feedback in voluntary contribution mechansims: An experiment on team production. *Res. Exp. Econ.* 5**2001**, *8*, 85–97.
23. Bigoni, M.; Suetens, S. Feedback and dynamics in public good experiments. *J. Econ. Behav. Org.* **2012**, *82*, 86–95.
24. Sell, J.; Wilson, R.K. Levels of Information and Contributions to Public Goods. *Soc. Forces* **1991**, *70*, 107–124.
25. Kreitmair, U. *Voluntary Disclosure of Contributions: An Experimental Study on Non-Mandatory Approaches for Improving Public Good Provision*; Working Paper; Indiana University: Bloomington, IN, USA, 2014.
26. Van der Heijden, E.; Moxnes, E. *Information Feedback in Public-Bads Games: A Cross-Country Experiment*; Working Paper; Tilburg University: Tilburg, Netherlands, 1999.
27. Carpenter, J. When in Rome: Conformity and the provision of public goods. *J. Socio-Econ.* **2004**, *33*, 395–408.
28. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
29. Croson, R.; Gneezy, U. Gender Differences in Preferences. *J. Econ. Lit.* **2009**, *47*, 448–474.
30. Sefton, M.; Shupp, R.; Walker, J. The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* **2007**, *45*, 671–690.
31. Frechette, G. Session-effects in the laboratory. *Exp. Econ.* **2012**, *15*, 485–498.
32. Samek, A.S.; Sheremeta, R. Recognizing contributors: An experiment on public goods. *Exp. Econ.* **2014**, *17*, 673–690.
33. Driscoll, J.; Kraay, A. Consistent covariance matrix estimation with spatially dependent panel data. *Rev. Econ. Stat.* **1998**, *80*, 549–560.
34. Vogelsang, T. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *J. Econ.* **2012**, *166*, 303–319.
35. Hoechle, D. Robust standard errors for panel regressions with cross-sectional dependence. *Stata J.* **2007**, *7*, 281–312.
36. Wilson, J.; Kelling, G. Broken Windows. *Atl. Mon.* **1982**, *249*, 29–38.
37. Keizer, K.; Lindenberg, S.; Steg, L. The Spreading of Disorder. *Science* **2008**, *322*, 1681–1685.
38. Kimbrough, E.O.; Vostroknutov, A. Norms Make Preferences Social. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2267135 (accessed on 22 September 2015).

Stable Sampling Equilibrium in Common Pool Resource Games

Juan Camilo Cárdenas, César Mantilla and Rajiv Sethi

Abstract: This paper reconsiders evidence from experimental common pool resource games from the perspective of a model of payoff sampling. Despite being parameter-free, the model is able to replicate some striking features of the data, including single-peaked frequency distributions, the persistent use of strictly dominated actionsand stable heterogeneity in choices. These properties can also be accurately replicated using logit quantal response equilibrium (QRE), but only by tuning the free parameter separately for separate games. When the QRE parameter is constrained to be the same across games, sampling equilibrium provides a superior fit to the data. We argue that these findings are likely to generalize to other complex games with multiple players and strategies.

Reprinted from *Games*. Cite as: Cárdenas, J.C.; Mantilla, C.; Sethi, R. Stable Sampling Equilibrium in Common Pool Resource Games. *Games* **2015**, *6*, 299–317.

1. Introduction

Common property is a regime involving a well-defined set of users, each of whom has rights to the extraction of an economically-valuable resource. Many inshore fisheries, grazing lands, forest areas and water resources are accessed under these conditions. The notion that such resources could be exploited in a manner that is sustainable and reasonably efficient is supported by numerous case studies, as well as theoretical models and experimental tests using subjects in the lab and the field. [1]

Experimental studies of common pool resource games have generally followed the protocol introduced by Ostrom *et al.* [3,4]. These experiments involve a finite number of rounds in which subjects take actions intended to mimic the extraction of a resource held as common property. This involves a negative externality, since higher levels of extraction reduce the marginal returns to the extractive effort of all players. Standard equilibrium analysis predicts over-extraction of the resource relative to efficient levels under these conditions.

The behavior of subjects in common pool resource experiments typically deviates substantially from this equilibrium prediction in a number of interesting

[1] See Ostrom [1] for theory and evidence, Bromley [2] for a collection of case studies, Ostrom *et al.* [3,4] for early experimental tests, Cárdenas [5] for more recent field experiments and Sethi and Somanathan [6] for an evolutionary argument.

respects. All available actions are chosen with a positive frequency, with strictly dominated actions being chosen persistently and often. The frequency distribution has a structure where levels of extraction that are less individually costly are selected more often. Average extraction is relatively stable over time, lying below equilibrium levels, but above efficient levels.

We argue in this paper that these patterns can be accurately replicated with a model of payoff sampling equilibrium (PSE) introduced by Osborne and Rubinstein [7], subject to a suitable refinement. The basic idea underlying this solution concept is that individuals try out multiple actions, observe payoffs and subsequently adopt actions that were the most rewarding. A sampling equilibrium is a distribution of actions in a population that reproduces itself, in the sense that the likelihood with which an action is selected under the sampling procedure matches the frequency with which it is currently being used.

While a PSE can involve the play of strictly dominated strategies with positive probability, it is also the case that every strict Nash equilibrium is also a sampling equilibrium, so the model does not predict that dominated actions must be played with positive probability. To obtain this stronger claim, one can use a refinement of sampling equilibrium based on dynamic stability [8]. The refinement selects a unique sampling equilibrium in the common pool resource games considered here, and this involves the play of dominated strategies with positive probability, as well as frequency distributions over actions very much like those observed in the data.

We compare sampling equilibrium with the more widely-used solution concept of quantal response equilibrium (QRE) introduced by McKelvey and Palfrey [9]. This is based on the idea that individuals make errors when responding to the behavior of others, have accurate beliefs about the distribution of opponent actions (taking full account of error rates) and best respond (with error) to these beliefs. In this paper, we focus on the logit QRE model, which has one free parameter to capture the rate at which errors are made. The model is extremely flexible and encompasses both Nash equilibrium (for zero error rates) and a uniform distribution over actions (for infinite error rates) as special cases. This flexibility allows for a very good fit to be obtained for any given set of experimental data by suitably tuning the free parameter [10]. However, the optimized parameter value can vary widely across treatments, even within the class of common pool resource games. This is especially the case if one treatment has an interior Nash equilibrium, while the other has a corner solution. If the QRE parameter is constrained to be equal across treatments, then payoff sampling provides a superior fit to the data.

We establish these claims by examining data from two experimental common pool resource games: the classic studies of Ostrom *et al.* [4] with interior equilibrium strategies and a more recent implementation by Cárdenas [11] with corner solutions. Both the parameter-free PSE and the single-parameter QRE models are capable

of producing a striking congruence between predicted and observed frequency distributions in both settings, but only if different, game-specific values of the error rate under QRE are allowed. In fact, the optimized value of the error rate parameter is more than two hundred-times greater for the Ostrom data than for the Cárdenas data. In the face of such a large difference across games with similar numbers of players and actions and nonlinear payoff functions of comparable complexity, it is hard to sustain the claim that the QRE parameter is merely capturing a degree of rationality among experimental subjects.

Of course, QRE and PSE are not the only solution concepts that have been proposed as alternatives to Nash equilibrium in the literature. In Section 5, we look at other alternatives, including action sampling, impulse balance, level-k reasoning and social preferences. Conditional on having a unique Nash equilibrium in pure strategies, as in standard common pool resource games, we show that action sampling simply reduces to the dominant strategy equilibrium. In addition, impulse balance equilibrium is not well-defined for these games. Level-k models predict that only Level-0 players choose dominated actions and that all dominated actions are chosen with equal frequency. None of these models can generate predictions that provide a qualitative match to the data. Social preferences do appear to be important, especially for non-student populations in Cárdenas [11], and hybrid models incorporating both social preferences and learning, along the lines of Arifovic and Ledyard [12], appear to be well worth developing. However, these contain a large number of free parameters, and it seems worthwhile to see how close a fit to the data can be obtained by a simple parameter-free solution concept, such as sampling equilibrium, especially in comparison to the widely-used one-parameter family of quantal response equilibria.

2. Experimental Common Pool Resource Games

The basic structure of a common pool resource game is as follows. There are n players, each of whom faces a set of ordered actions $A = \{a_1, ..., a_m\}$, interpreted as feasible levels of resource extraction. It is typically assumed that a_1 is either zero or one. Letting $x_i \in A$ denote the extraction of player i, the aggregate extraction level is:

$$X = \sum_{j=1}^{n} x_j$$

The payoffs to each player depend only on her own action and the aggregate action and may be written:

$$\pi_i = g(x_i, X)$$

The function g is increasing in the first argument and decreasing in the second. That is, given a level of aggregate extraction, those with higher individual extraction levels get higher payoffs, but an increase in one person's extraction level lowers the payoff of all others. This external effect results in a divergence between individual incentives and collective interests, as well as equilibria with inefficiently high levels of extraction under standard assumptions. The stage game is repeated for a fixed number of rounds.

Various versions of this game have been studied experimentally, following the pioneering work of [3,4]. We shall focus on results from two implementations. In Ostrom *et al.* [4], eight participants, each endowed with 10 tokens, had to divide these between two markets: one with a fixed rate of return and another with a rate dependent on the total amount invested by the group. Here, $n = 8$, and $A = \{0, ..., 10\}$ is the set of possible investments in the latter market. The payoff to individual i was given by:

$$\pi_i = w(e - x_i) + \frac{x_i}{X} f(X) \tag{1}$$

if $X > 0$ and $\pi_i(x) = we$ otherwise, where $e = 10$, $w = 0.05$ and the function f, representing total output from the common pool resource, was given by:

$$f(X) = 0.23X - 0.0025X^2$$

The interpretation is that each of the $e - x_i$ tokens invested in the first market earns a fixed rate w, while each of the x_i tokens invested in the second market (or common pool) earns an equal share of aggregate output. The unique Nash equilibrium action of this game is $x_i = 8$, resulting in aggregate extraction $X = 64$. Aggregate payoffs are maximized if $X = 36$, which requires an average extraction level below five. Hence, both equilibrium and efficient extraction levels are interior.

The experiments involved 56 participants divided into seven groups. Subjects were shown in tabular form the total output $f(X)$, as well as the per token output $f(X)/X$ for various values of X. In each group, subjects interacted initially for 10 rounds without any kind of communication, receiving feedback only about the aggregate extraction level after each round.

Now, consider the implementation by Cárdenas [11], where $n = 5$, $A = \{1, ..., 8\}$, and the payoffs are given by:

$$\pi_i = ax_i - \frac{bx_i^2}{2} + \phi(ne - X) \tag{2}$$

where a, b, and ϕ are positive constants and $e = 8$ is the maximum extraction level.

As long as $n\phi > a$, aggregate payoffs are maximized when all individuals i choose the lowest extraction level $x_i = 1$, and any symmetric action profile other

than this is Pareto-dominated. In addition, if $a > \phi$ and b is sufficiently small, then equilibrium extraction levels will be inefficiently high. Experimental parameters were set as follows: $a = 60$, $b = 5$ and $\phi = 20$. Under these conditions, the choice of $x_i = 8$ is a dominant strategy, while aggregate payoffs are maximized if $x_i = 1$ is chosen by all players. The social dilemma then appears in its starkest form: efficiency demands minimum extraction, while maximum extraction is a dominant strategy.

As in Ostrom *et al.* [4], individuals were provided information about the payoff structure in tabular form: given their own extraction and the aggregate extraction of all others, they could see the resulting payoff that they would receive. They chose extraction levels simultaneously in each of ten rounds. At the end of each round, once all choices had been made, the aggregate extraction level was computed and announced publicly. This was the only information received by the subjects after each round. The experiment involved 230 participants drawn from a population of students and divided into 46 groups [2].

There are five aspects of behavior that emerge consistently across both implementations: almost all available actions are chosen with non-negligible frequency in the subject population as a whole; more individually-costly actions tend to be chosen with lower frequency; most subjects choose a variety of strategies over several rounds; the average level of extraction within groups is stable over time and intermediate between efficient and equilibrium levels; and the heterogeneity in action choice does not appear to diminish over the rounds.

The left panel of Figure 1 displays the distribution of strategies aggregated by participants and rounds for the two experimental studies. In general, higher extraction levels are chosen with greater frequency. The maximum extraction is the modal choice in both settings, although it is the dominant action only in Cárdenas [11]. These aggregates mask considerable variation in action choice at the individual level across rounds: in Ostrom *et al.* [4], 73% of subjects chose the Nash equilibrium action at least once, while nobody chose this action in every round. In Cárdenas [11], 70% of subjects chose maximum extraction at least once, while less than 1% chose this action in every round.

---

[2] Both sets of experiments had a second stage designed to uncover the effects of different institutions: one-shot and repeated communication, pecuniary and non-pecuniary penalties and endogenous rule-making. As we show in Table 1, there is remarkable consistency across the two studies in the first-stage data, which is the basis of our analysis here. In addition, Cárdenas [11] also reports results from a population of villagers drawn from communities in which management of resources held as common property is a routine and economically-important activity. Here, we focus on results from the student population, which facilitates comparison to Ostrom *et al.* [4]; we briefly discuss the field data in Section 5. Other experiments with similar payoff functions and results are also reported in Rodriguez-Sickert *et al.* [13] and Cárdenas [5].

The right panel of Figure 1 shows the distribution of actions for all participants in their last three rounds of play (Rounds 8–10). Higher extraction levels continue to be chosen with greater frequency than lower extraction levels on the whole, and there is a slight increase in the frequency of the modal choice.

Table 1 summarizes some of the key features of the two designs, including Nash equilibrium and joint surplus maximizing per-capita extraction levels. It also provides some descriptive statistics regarding the level and variability of extraction across rounds. The considerable variability across rounds in actions chosen by given individuals is revealed in Part (b) of the table.

Individuals sample, on average, between four and five different strategies. In Ostrom *et al.* [4], 7% of individuals limited themselves to a single action, and about 70% of them chose at least four distinct actions across the ten rounds. In Cárdenas [11], less than 1% of individuals sampled a single action, while 88% of them chose no less than four different actions. This suggests considerable experimentation on the part of subjects.

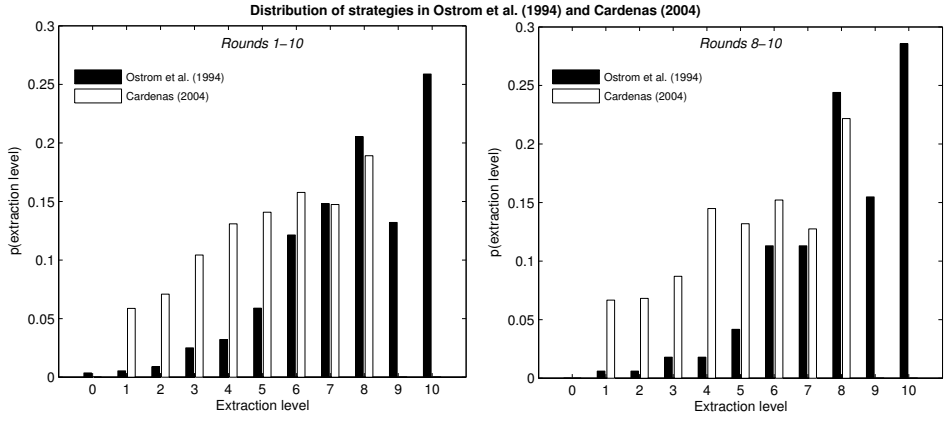


Figure 1. Distribution of strategies in Ostrom *et al.* (1994) and Cárdenas (2004).

Next, consider the manner in which the average action choice varies over rounds, reported in Part (c) of Table 1. We see neither a convergence to the equilibrium action, nor to the efficient action in any of the experimental studies. In fact, the average extraction level remains quite steady at an intermediate level.

Part (d) of the table shows that the within-group variance does not seem to converge to zero or even to diminish systematically over time in either of the two studies. We observe that the average standard deviation per group per round oscillates around two and does not exhibit a trend across stages of the game. This stability in the variability of actions within groups is evidence against the hypothesis that individuals coordinate on a pure-strategy Nash equilibrium of a different game,

obtained by transforming the material payoffs to allow for social preferences or reciprocity norms. For instance, if social preferences were to transform a social dilemma into a coordination game with multiple equilibria, convergence over time to one of these would result in the variance of within-group actions to decline over successive rounds.

Table 1. Summary statistics for two experimental studies.

	Cárdenas [11]	**Ostrom *et al.*** [4]
a.Experimental setting		
Number of subjects	230	56
Subjects per group	5	8
Number of rounds	10	10
Action set	{1,. . . ,8}	{0,. . . ,10}
Nash equilibrium	8	8
Surplus maximizing per-capita extraction	1	4.5
b. Strategies sampled in ten rounds		
Median	6	4
Mean	5.37	4.16
Standard deviation	1.53	1.60
c. Average extraction levels		
Rounds 1–5	5.16	7.56
Rounds 6–10	5.31	7.86
d. Within group standard deviation per round		
Rounds 1–5	1.96	2.11
Rounds 6–10	1.94	1.93
Rounds 11–20 (subsample)	2.12	2.02

To rule out the possibility that ten rounds are too few for convergence to be obtained, the last row of the table shows the average within-group standard deviation for the subsample of sessions in which interaction occurred over twenty rounds rather than ten. Again, we see that that the variance of actions is maintained at a roughly constant level throughout the additional ten rounds.

The distribution of actions chosen by participants and the stability over time of both the mean and the variance of this distribution suggest that conventional models based on social preferences cannot fully account for the data. In particular, the heterogeneity of actions at the level of an individual is suggestive of sampling and action selection in response to observed payoffs as in payoff sampling equilibrium or, alternatively, to the incorporation of noisy behavior into the best response function as in quantal response equilibrium.

We next show that both PSE and QRE can provide a good fit to the data. Both models predict the play of strictly dominated strategies in equilibrium, though for very different reasons. In QRE, dominated actions are played by mistake, while

in PSE, they are played because they have a positive probability of being the most rewarding under the sampling procedure.

3. Payoff Sampling and Quantal Response

The key idea underlying the concept of payoff sampling equilibrium is that when faced with a novel strategic situation, individuals will experiment with actions, observe outcomes and choose among available choices based on realized payoffs. The problem of selecting an action is viewed, in effect, as a multi-armed bandit problem in which a given probability distribution over the actions of others determines the consequences of any given choice by the subject in question. A sampling equilibrium is a distribution that is self-generating, in the sense that the probability of selecting an action conditional on this distribution matches the probability with which this action is taken in the distribution.

Formally, consider a symmetric game with actions $\{a_1, ..., a_m\}$, and let $p = (p_1, ..., p_m)$ denote some arbitrary probability distribution over these actions. Now, suppose that a player samples each of the actions exactly once, and on each such trial, her opponent plays the mixed strategy given by p. Once all actions have been sampled, each will be associated with a realized payoff. Suppose that the player who is sampling selects the action with the highest realized payoff, breaking ties with uniform probability. Let $w_i(p)$ denote the probability that action a_i is selected under this procedure. One interpretation of $w_i(p)$ is that it denotes the probability that action a_i is selected when one's opponent at each step of the sampling procedure is independently drawn from a large population in which a proportion p_i always plays a_i.

A sampling equilibrium is simply a mixed strategy p^* that satisfies:

$$w_i(p^*) = p_i^*$$

for all i. That is, for each action a_i, the likelihood that it will be selected is equal to the likelihood with which it is currently being played. This may be interpreted as a steady state of a dynamic process in which a large incumbent population is choosing actions in the proportions $p_1, ..., p_n$, while new entrants choose actions in accordance with $w_1(p), ..., w_n(p)$. Such a dynamic process is implicit in the justification for sampling equilibrium offered by Osborne and Rubinstein and requires that the rates of change $\dot{p}_i$ satisfy:

$$\dot{p}_i > 0 \text{ if and only if } w_i(p) > p_i \tag{3}$$

Stability with respect to this dynamic process can be used as a criterion for selection among multiple sampling equilibria.

The set of sampling equilibria can be large. It is easily seen, for instance, that every strict Nash equilibrium is also a sampling equilibrium. However, sampling equilibria can also have surprising features and involve the choice of actions that are strictly dominated. In fact, this can occur even in sampling equilibria that are stable under the dynamics (3).

In a symmetric n-player game with more than two players, a sufficient condition for the instability of a symmetric sampling equilibrium is that the corresponding action profile satisfies a condition called inferiority [8]. Specifically, a symmetric action profile $(a_q, \ldots, a_q)$ is said to be inferior if, when $n-2$ players continue to choose action a_q while one player deviates to some action $a_i \neq a_q$, then for the remaining player, there exists at least one response $a_j \neq a_q$, such that, for this player, the resulting payoff is strictly preferred to the outcome at $(a_q, a_q, \ldots, a_q)$.

To see how a dominant strategy equilibrium profile can be inferior (and hence unstable under the sampling dynamics), consider a simple three-player, two-action public goods game in which each player has an indivisible unit endowment and can either keep it or contribute it to the provision of a public good. If the total contribution is $z \in \{0, 1, 2, 3\}$, then a player contributing $y \in \{0, 1\}$ to the public good obtains payoff $1 - y + 2z/3$. Clearly, a dominant strategy for each player is to contribute nothing, in which case the payoff to all players is one. However, this action profile is inferior: if one player contributes while another does not, the third player gets $4/3 > 1$ if she also contributes. Hence, the dominant strategy equilibrium is unstable under the dynamics (3) by Theorem 1 in [8], and strictly dominated strategies must be played with positive probability in any stable sampling equilibrium of the game [3].

It is easily verified that for the payoff function (2) and the parameter values chosen by Cárdenas [11], the dominant strategy equilibrium in which all players choose the maximum extraction level $x_i = 8$ is inferior and therefore unstable. For this action profile, all five players receive a payoff of 320. If three of the five continue to choose the dominant strategy while one switches to any other extraction level, the remaining player can guarantee himself at least 337 by choosing an action other than the dominant strategy. Of course, the payoff from sticking to the dominant strategy would be still greater, but the inferiority condition is nevertheless satisfied. Hence, the dominant strategy equilibrium is unstable, and the only stable sampling equilibrium involves the playing of dominated strategies with positive frequency. The precise frequencies with which the various actions are chosen in the unique stable sampling equilibrium for this game, as well as that of Ostrom *et al.* [4] are described in more detail below.

---

[3] Inferiority is not sufficient for instability in two-player games, and in fact, the dominant strategy equilibrium in the prisoners' dilemma is stable under the sampling dynamics.

Next, consider quantal response equilibrium. Here, the key idea is that subjects choose the strategy with the highest utility with a probability that depends on the utility difference with respect to alternative strategies [14]. It is assumed that the error rate (*i.e.*, the chance of picking a suboptimal strategy) is common knowledge and is therefore accounted for by each subject in forming his best response. Consider again a symmetric game with actions $\{a_1, ..., a_m\}$ with a probability distribution $p = (p_1, ..., p_m)$ over actions. Letting $E_k(p)$ denote the expected payoff when playing action a_k when faced with p, the logit QRE is given by the following set of $m - 1$ equations:

$$p_k = \frac{e^{\lambda E_k(p)}}{\sum_{k=1}^{K} e^{\lambda E_k(p)}} \tag{4}$$

The parameter λ varies inversely with the error rate and may be interpreted as the degree of player rationality. When $\lambda \to \infty$ (the error rate tends to zero), subjects are unboundedly rational, and logit QRE is equivalent to Nash equilibrium. When $\lambda \to 0$ (the error rate is very large), subjects are acting randomly, and in equilibrium, each strategy is chosen with the same probability regardless of its expected payoff. A positive probability of play for dominated strategies is therefore guaranteed as long as we assumed a bounded degree of rationality captured through a finite value for λ.

For the experiments considered here, the predictions for the two-solution concepts (stable PSE and QRE) are shown in Table 2. The λ parameter for the QRE calculation corresponds to the value that minimizes the mean squared error for each CPRgame, *i.e.*, the inverse of the error rate that best fits the distribution of optimal and suboptimal choices. The λ parameter was computed separately for each game using Equations (1) and (2). The optimal λ values are 0.09 and 18.2 for the reported data in Cárdenas [11] and Ostrom *et al.* [4], respectively [4].

The comparison is also shown graphically in Figure 2. For Cárdenas [11], the mean squared errors for the QRE and the PSE models are 0.0001 and 0.0111, respectively. For Ostrom *et al.* [4], where the Nash equilibrium is interior, the mean squared errors are 0.0091 and 0.0105 for QRE and PSE, respectively. Allowing a different estimated parameter λ for each experimental setting, QRE outperforms PSE. The difference in their mean squared errors is especially stark in games where the Nash equilibrium is a corner solution, but the dominance of QRE over PSE is less clear if the individually-rational strategy is an interior point of the strategy set. Intuitively, the logistic QRE is very precise in the calibration of a straight line, as occurs in the Cárdenas [11] dataset, because the maximum extraction level allowed

[4] By way of comparison, Brunner *et al.* [15] propose a value of $\lambda = 1.05$ for fitting the QRE model to data from a variety of 2×2 games originally reported in Selten and Chmura [16].

matches the Nash equilibrium. By definition, strategies closer to the optimal strategy are played with greater probability under QRE. Therefore, the predicted probabilities are strictly increasing. For the dataset from Ostrom *et al.* [4], QRE and PSE perform about equally well. The interior Nash equilibrium substantially lowers the predictive capacity of QRE.

Table 2. Computed equilibria for the CPRgames.

	Cárdenas [11]										
	x = 0	x = 1	x = 2	x = 3	x = 4	x = 5	x = 6	x = 7	x = 8	x = 9	x = 10
Nash	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	-	-
PSE	-	0.015	0.033	0.065	0.117	0.146	0.185	0.201	0.238	-	-
QRE (λ = 0.09)	-	0.058	0.077	0.099	0.121	0.142	0.159	0.170	0.173	-	-
Observed	-	0.059	0.071	0.104	0.131	0.141	0.158	0.147	0.189	-	-
	Ostrom *et al.* [4]										
	x = 0	x = 1	x = 2	x = 3	x = 4	x = 5	x = 6	x = 7	x = 8	x = 9	x = 10
Nash	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
PSE	0.000	0.000	0.00001	0.003	0.023	0.063	0.117	0.165	0.194	0.219	0.216
QRE (λ = 18.2)	0.003	0.007	0.014	0.027	0.047	0.075	0.109	0.145	0.177	0.196	0.199
Observed	0.004	0.005	0.009	0.025	0.032	0.059	0.121	0.148	0.205	0.132	0.259

Our claim is that the large difference in the implied value of λ is not due primarily to differences in the rationality of subjects or the complexity of the two games, but rather to the increasing difficulty of calibration when the action profile has an interior mode. The experimental pools are comparable, as both experiments were conducted with college students (although separated across time and space). The key difference between the games is the location of the socially-desirable and the individually-rational extraction levels within the choice set [5]. It is striking, therefore, that the difference in implied values of λ reaches two orders of magnitude, as we move from 0.09–18.2, for games that are structurally similar and have a comparable number of players.

An alternative approach to calibration, as used in Selten and Chmura [16], is to define a unique value for λ that minimizes the joint mean squared errors. As is shown in the continuous bold line in Figure 3, this value of λ is 0.09, as in the Cárdenas [11] setting. The intuition behind this result is that QRE has a greater goodness of fit for the Cárdenas [11] sample for low values of λ, about one order of magnitude lower than the mean squared errors for Ostrom *et al.* [4], but it rapidly decreases its predictive success as λ increases. Using a single value for λ, the minimum mean squared error for QRE is 0.0784, much larger than the 0.0216 obtained for PSE. This

[5] Carpenter and Cardenas [17] use a CPR experimental design with interior solutions much like Ostrom *et al.* [4], with similar results.

difference is also observable in Figure 3 by comparing the black continuous curve with the gray straight line.

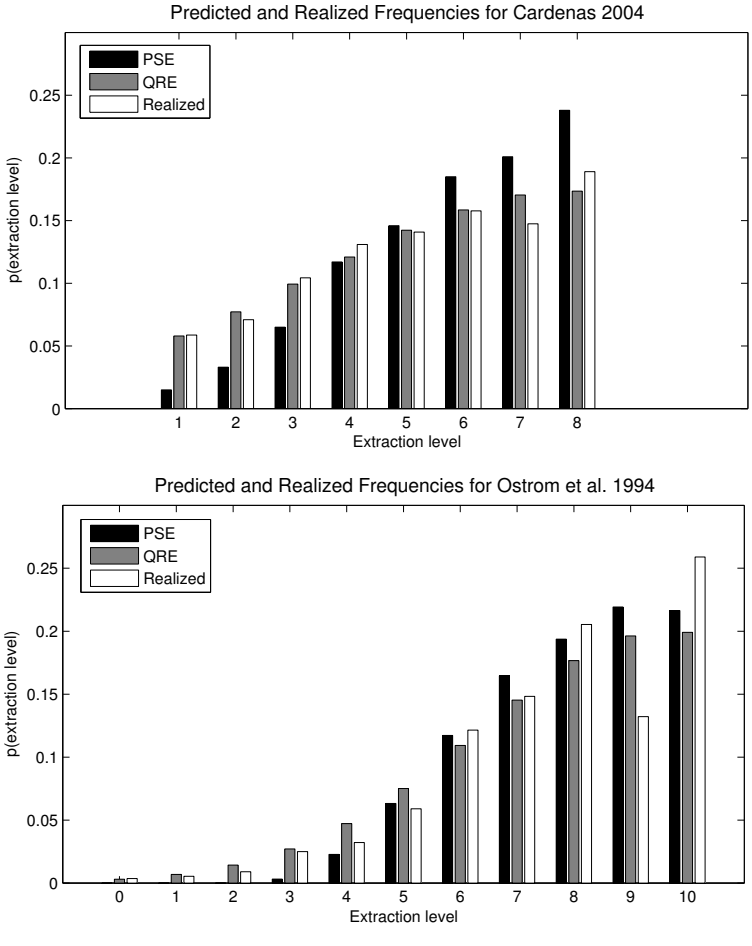


Figure 2. Comparison between payoff sampling equilibrium (PSE), quantal response equilibrium (QRE) and experimental data.

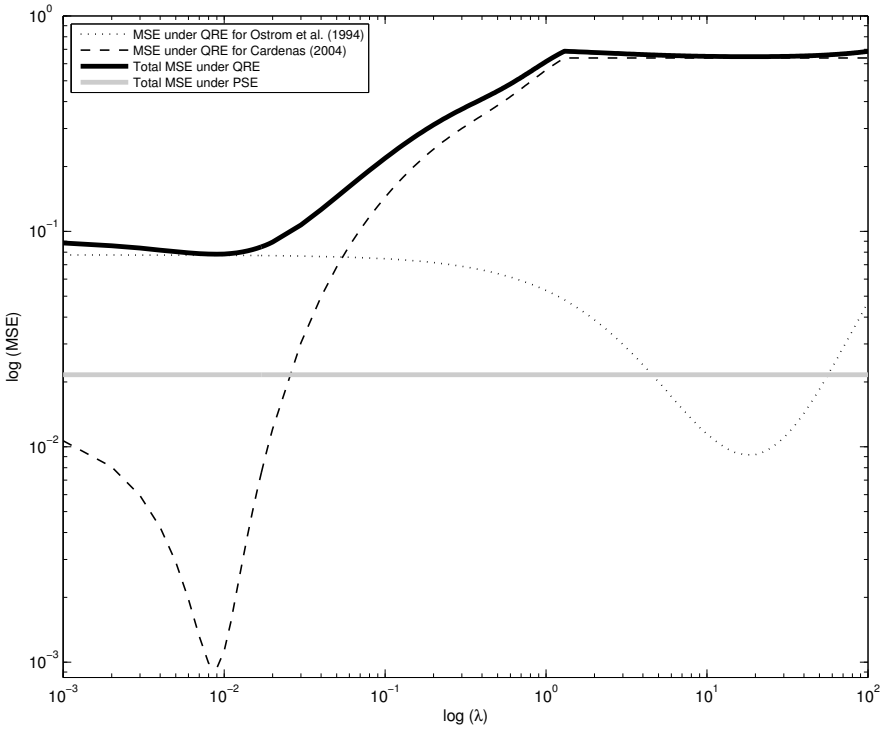


Figure 3. Aggregated mean squared errors as a function of λ.

4. Statistical Tests

The visual correspondence between observed and predicted extraction levels is striking for the PSE and QRE models, but falls short of a statistical test of the model. We now turn to a more formal analysis of the match between theory and data.

Table 3 reports results from a regression of observed *versus* predicted frequencies of use for each action (both expressed as percentages). Of the m observations, one per available action, only $m - 1$ are independently distributed. Therefore, we use a bootstrap procedure in which $m - 1$ observations are randomly drawn in each one of the one hundred repetitions of the ordinary least squares estimation. Reported standard errors are corrected for the bootstrap procedure. Standard errors are larger than in the OLS regression with m observations, and the coefficients only differ after the third decimal number (regression results not shown but available upon request). Here, the prediction is a slope not different from unity, and lower slopes are indicative of a bias towards efficiency. We easily reject the hypothesis that the slopes equal zero for the two experiments under both equilibrium concepts. Furthermore,

we cannot reject the hypothesis that the slope is one for the Cárdenas [11] data fitted with the QRE, but we clearly reject it for the PSE prediction.

For the Ostrom *et al.* [4] data, the relative goodness of fit is reversed. The slope coefficient for the PSE does not statistically differ from one, while for the QRE, we observe a slope considerably larger than unity along with a negative intercept. The reason for this poor calibration is that the low λ value that minimizes the mean squared error for the pooled data provides an exceptionally good fit for the game with the corner solution, at the cost of a very poor fit for the game with the interior solution.

Table 3. OLS: observed *versus* predicted percentage frequencies.

	Cárdenas [11]		**Ostrom *et al.*** [4]		
	PSE	QRE ($\lambda = 0.09$)	PSE	QRE ($\lambda = 0.09$)	QRE ($\lambda = 18.2$)
Slope (β_1)	0.528 ***	0.982 ***	0.887 ***	44.06 ***	1.068 ***
	(0.0705)	(0.165)	(0.148)	(9.909)	(0.174)
Constant	5.901 ***	0.234	1.024*	−391.5 ***	−0.611
	(0.947)	(2.019)	(0.570)	(89.98)	(0.876)
Ho: $\beta_1 = 1$	101.76 ***	0.03	0.54	27.86 ***	0.15
	(0.0001)	(0.8737)	(0.4804)	(0.0005)	(0.7103)
Observations	8	8	11	11	11
R-squared	0.950	0.932	0.879	0.817	0.886

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

To illustrate the dependence of the QRE goodness of fit on its parameter value, consider the regression estimates when the value of λ equals 18.2, which minimizes the mean squared errors for the Ostrom *et al.* [4] data. As shown in the last column of Table 3, if this is the case, we cannot reject the hypothesis that the slope equals one.

We propose an additional test for the goodness of fit of PSE and QRE. We separately regress observed and predicted percentage frequencies on the extraction level using a quadratic specification to allow for nonlinearities. This allows us to compare the estimated coefficients for the observed data to the estimated coefficients for the predicted frequencies. The bootstrap procedure is also applied in this estimation by randomly drawing $m - 1$ observations in each repetition. The results are shown in Table 4.

A Hausman test may be used to compare the regression coefficients, based on the null hypothesis that coefficients are the same across the two specifications. The Hausman statistic will be insignificant if the observed and predicted frequencies are sufficiently similar. We fail to reject the hypothesis of identical coefficients for the comparison between Cárdenas [11] data and predicted QRE, but we reject it for the PSE. For the Ostrom *et al.* [4] data, the roles are reversed yet again. We reject the hypothesis that the coefficients are structurally similar for the observed data and the

QRE with $\lambda = 0.09$, but we fail to reject it when comparing the observed data with the PSE. Nevertheless, if we set $\lambda = 18.2$, we cannot reject the hypothesis that the polynomial calibration with the observed frequencies and the QRE are structurally the same.

Table 4. OLS: observed/predicted percentage *versus* the second order polynomial for extraction.

	Cárdenas [11]			**Ostrom *et al.*** [4]			
	Observed	PSE	QRE ($\lambda = 0.09$)	Observed	PSE	QRE ($\lambda = 0.09$)	QRE ($\lambda = 18.2$)
Extraction	2.776	3.643 **	2.957 ***	0.669	0.983	0.0731 ***	1.278
	(1.919)	(1.541)	(0.676)	(1.850)	(1.979)	(0.000)	(1.111)
Extraction^2	−0.115	−0.0357	−0.134 *	0.179	0.170	−0.00189 ***	0.102
	(0.206)	(0.160)	(0.071)	(0.174)	(0.191)	(0.0000)	(0.107)
Constant	2.930	−2.982	2.598	−0.532	−1.789	8.791 ***	−0.864
	(4.420)	(3.487)	(1.619)	(4.471)	(4.914)	(0.0006)	(2.737)
Hausman test		47.91 ***	0.01		1.54	18.12 ***	0.45
		(0.0000)	(0.9949)		(0.4619)	(0.0001)	(0.7989)
Observations	8	8	8	11	11	11	11
R-squared	0.948	0.990	0.994	0.894	0.944	1.000	0.974

The reported χ^2 statistic for each Hausman test corresponds to the structural comparison between observed; data and the relevant equilibrium prediction listed at the top of the column. Standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

To summarize, the statistical tests confirm what visual inspection led us to believe: the concepts of stable sampling equilibrium and quantal response equilibrium can explain a great deal of the variation in frequencies from the experimental data. We also observe that QRE fits better than PSE for the game with a corner Nash equilibrium, while the opposite is true when the individually-rational strategy is an interior point in the strategy set.

After exploring how PSE and QRE fit the experimental data separately, our next step is to compare the relative predictive success of the two models. Following the comparison made in Selten and Chmura [16], we apply the Wilcoxon-matched pairs signed rank test to the mean squared errors of the two equilibrium models. For the computation in the QRE, we use $\lambda = 0.09$, a value that minimizes the mean squared errors of the whole sample, *i.e.*, the two experimental settings. We consider each group as an independent observation, having a total of 46 observations for the data in Cárdenas [11] and seven observations for the data in Ostrom *et al.* [4].

Results of the statistical tests are displayed in Table 5. We do not find any differences in the predictive success for the whole sample. Nevertheless, if we performed the Wilcoxon signed rank test separately for the two experimental settings, the results are different. Moreover, they confirm what we have been discussing in this section: for the game with the extreme Nash equilibrium, from

Cárdenas [11], QRE has greater predictive power than PSE, and the test is significant at the five percent level. For the game with the internal Nash equilibrium, Ostrom *et al.* [4], the larger predictive power of PSE is confirmed by this test. The statistical significance at the five percent level is particularly surprising given the limited number of available observations.

Table 5. Wilcoxon signed rank test.

	Experimental Datasets		
	Pooled	**Cárdenas** [11]	**Ostrom *et al.*** [4]
Independent groups	53	46	7
z	0.580	2.540 **	−2.366 **
p-value	0.562	0.011	0.018

$z > 0$ indicates a better fit from QRE. $z < 0$ indicates a better fit from PSE. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

5. Alternative Models

We have focused to this point on two alternatives to Nash equilibrium: payoff sampling and quantal response. There are, of course, other models of behavior in experimental games. These can be divided into two groups: those that look for departures from Nash equilibrium based on learning or reasoning and those that maintain the Nash hypothesis, but allow for the possibility that material payoffs do not define the game that individuals are actually playing, for instance because of social preferences or reciprocity norms. Both of these approaches have given rise to substantial literature in their own right, as well as the recent development of hybrid models.

In an analysis of data from multiple 2×2 games, Selten and Chmura [16] consider four alternatives to Nash equilibrium [6]. In addition to the two discussed extensively above, they consider action sampling equilibrium (ASE) and impulse balance equilibrium (IBE). Under action sampling, individuals take a random sample of observations of their opponent's strategies and then optimize against this sample. Under impulse balance, it is assumed that, being informed of his realized and foregone payoffs, the subject could experience downward or upward impulses after comparing the realized payoff with his "natural aspiration level". Under this equilibrium concept, impulses towards one strategy must be equally compensated by impulses towards the other strategy. Chmura *et al.* [18] extend their comparison between impulse balance equilibrium and Nash equilibrium to 3×3 games with

[6] In Selten and Chmura [16], payoff sampling equilibrium is described as a one-parameter model, since they allow for different sample sizes and use this as a tunable parameter to obtain the best fit. Our version of the model is clearly parameter-free.

completely mixed and partially mixed equilibria and find that generalized impulse balance performs best in predicting the experimental data.

When these concepts are generalized to allow for multiple players and actions and applied to the common pool resource game, action sampling simply reduces to the dominant strategy equilibrium, while impulse balance is not well defined. The impulse balance equilibrium can be computed if strategies are dominated by a mixture of two pure strategies, as in Chmura *et al.* [18], but not when they are dominated by a pure strategy, as in our CPR games. For games with dominant strategies, impulse balance equilibrium cannot be calculated, because, by construction, all of the downward impulses are zero for the dominant strategy.

To see why action sampling reduces to the dominant strategy equilibrium, recall that a player takes a random sample of m observations of his counterpart's strategies and applies a best response to this sample. Since the effect of effort levels is additive in common pool resource games, the only payoff relevant attribute of the sample is the aggregate extraction $X_{-i} = \sum_{j \neq i} x_j$ of other players, and the best response to any given sample is therefore always the dominant strategy.

An alternative behavioral model in which dominated strategies might be played with positive probability is level-k reasoning [19,20]. This model assumes a heterogeneous population in terms of their degree of iterative reasoning k. Choices from Level-0 players are assumed to be randomly chosen, *i.e.*, they are drawn from the strategy set with the same probability. Level-1 players best respond to Level-0, Level-2 to Level-1, and so on. Let us define α as the fraction of Level-0 subjects and $1 - \alpha$ as the fraction of all level-k subjects with $k \geq 1$. In a game with J pure strategies, the predicted proportion of play will be α/J for all dominated strategies and $1 - \alpha(1 - J)/J$ for the dominant strategy. This model fails to predict that less costly strategies are played more often even if they are strictly dominated, as is shown in Figure 1.

A different strand of the literature on models of behavior in experimental games develops the idea that the material payoffs faced by experimental subjects do not accurately reflect their objective functions, because they care about the entire payoff distribution and, perhaps, also because their utilities are belief dependent, as in the theory of psychological games [7]. While the persistent variability in actions across rounds in Table 1 is evidence against coordination on a pure-strategy Nash equilibrium of a transformed game, it is nevertheless likely that social preferences

---

[7] Specific forms of other-regarding preferences capturing concerns for fairness and reciprocity have been proposed by Levine [21], Fehr and Schmidt [22], Bolton and Ockenfels [23], Charness and Rabin [24] and Cox *et al.* [25]. Belief-dependent utilities play a central role in Rabin [26], Dufwenberg and Kirchsteiger [27] and Falk and Fischbacher [28], all of whom employ the theoretical construct of psychological games [29].

and norms do play a significant role in accounting for the choices of subjects. This is especially the case for results from field experiments. In addition to the experiment with student populations, Cárdenas [11] also conducted a number of trials with villagers for whom common pool resource management is a routine activity. These subjects were less likely to choose the dominant strategy (15%, rather than 19% of the time) and more likely to choose efficient extraction (12% of the time, more than twice as often as the students). One-third of villagers never chose the dominant strategy in any round.

For this population of villagers, the mean squared errors for the QRE (with $\lambda = 0.09$) and PSE models are 0.0145 and 0.0473, respectively. These values are considerably larger than those derived from the student sample, for which mean squared errors were 0.0001 and 0.0111, respectively. The error for the sample of villagers drops to 0.0011 for the QRE model if the λ parameter is recalibrated to 0.000. The increase in predictive power for QRE here comes at the cost of assuming a very large error rate or, equivalently, null rationality among the villagers. A more plausible interpretation is that social preferences or norms of restraint play a greater role among villagers than they do among the pool of urban students, mitigating the effects of sampling dynamics. Put differently, any explanation of behavior based on sampling needs to take into account the fact that subjects have other-regarding preferences and that these may differ systematically across different populations.

These two distinct approaches (social preferences and learning) have been merged in recent work by Arifovic and Ledyard [12], who embed social preferences in a learning model with the goal of accounting for the basic stylized facts emerging from public goods experiments. Social preferences are introduced by allowing for individuals to care about both efficiency and envy, where the former is represented by average payoffs and the latter by any shortfall in one's own payoff relative to the average. Learning is modeled as follows. At each period, there is a finite list of remembered actions from which choices are made in accordance with a probability measure. The items on this list need not be unique: a given action can have multiple replicates. Replication in each period takes place on the basis of hypothetical performance in the immediately preceding period, with an action that would have performed poorly being replaced by one that would have done better. This measure of hypothetical performance also determines which action is played in any given period, with higher valued actions being selected with greater likelihood. In addition, the list of remembered actions itself changes over time, as incumbent items are replaced by novel ones with some probability.

The Arifovic and Ledyard model has six free parameters, while sampling equilibrium has none. This makes a meaningful performance comparison between the two approaches difficult, and we have accordingly focused on the more parsimonious quantal response model instead. Here, we find sampling to be very

competitive, especially if the free parameter for quantal response is constrained to be the same across experiments with a similar payoff structure and degree of complexity. More generally, the development of comprehensive models with many free parameters incorporating a variety of psychological effects alongside the exploration of simple, parsimonious models with few if any free parameters seems to be worthwhile.

6. Conclusions

Observed behavior in laboratory experiments under a common-pool resource dilemma shows that individuals often choose strictly dominated strategies even after a number of repeated rounds. This behavior yields outcomes that deviate systematically from the Nash equilibrium and allow these groups to mitigate the tragedy of the commons.

We have argued in this paper that a simple model of behavior based on the dynamics of sampling can be used to account for these experimental regularities. Even without incorporating other-regarding preferences, qualitative predictions based on stable sampling equilibrium match observed patterns remarkably well. In particular, the predicted lower frequencies for strategies that are more individually costly appear as a consistent and striking feature of the data when one restricts attention to the typical laboratory subject pools.

A distinctive feature of the sampling equilibrium concept is that it is parameter-free. The one-parameter family of quantal response equilibria can outperform sampling equilibrium in individual treatments, especially those involving corner solutions. However, when the error-rate parameter is constrained to be the same across treatments and games with interior solutions are also considered, then sampling equilibrium provides a somewhat superior fit.

We do not claim that the concept of stable sampling equilibrium alone can account for the experimental findings; it is clear that social preferences and norms also play a role, especially for populations with a high dependence on common property, close proximity and enduring social ties. Our purpose, rather, has been to bring the attention of experimental researchers to a versatile theoretical construct that might be very effective in accounting for observed data when used in conjunction with other approaches.

We suspect that the effectiveness of the sampling equilibrium concept in the common pool resource environment arises from the complexity of the game: there are multiple players; many actions for each of them; and the mapping from action profiles to payoffs is far from transparent. Under these conditions, it makes sense to explore options and to settle on those that happen to produce favorable results.

Merging a theory of social preferences with sampling dynamics seems both tractable and worthwhile, although the resulting model will clearly not be

parameter-free. Furthermore, the concept of stable sampling dynamics is versatile and can be applied to any normal form game. Testing predictions based on this hypothesis in other settings is accordingly an interesting and potentially fruitful area for future research.

Acknowledgments: For helpful comments on earlier versions, we thank Emily Breza, Thorsten Chmura and John Ledyard, as well as participants at the following seminars and conferences: Seminario CEDE (Colombia); Social Dilemmas 2013 (CalTech); Antigua Experimental Workshop (Guatemala); International Conference of the Courant Research Centre Poverty, Equity, and Growth in Developing and Transition Countries and the Ibero-America Institute of Economic Research 2014 (Gottingen); London Experimental Workshop 2014. The second author gratefully acknowledges support through ANR-Labex IAST.

Author Contributions: All of the authors contributed equally to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ostrom, E. Governing the Commons: The evolution of institutions for collective action. *Am. Political Sci. Rev.* **1990**, *92*, 1–22.
2. Bromley, W. The Commons, Property and Common Property Regimes. In *Making the Commons Works Theory, Practice and Policy*; Bromley, W., Ed.; Institute of Contemporary Studies: San Francisco, CA, USA, 1992; pp. 3–15.
3. Ostrom, E.; Walker, J.; Gardner, R. Covenants With and Without a Sword: Self-Governance is Possible. *Am. Political Sci. Rev.* **1992**, *86*, 404–417.
4. Ostrom, E.; Gardner, R.; Walker, J. *Rules, Games and Common-Pool Resources*; University of Michigan Press: Ann Arbor, MI, USA, 1994.
5. Cárdenas, J.C. Social Norms and Behavior in the Local Commons as Seen Through the Lens of Field Experiments. *Environ. Res. Econ.* **2011**, *48*, 451–485.
6. Sethi, R.; Somanathan, E. The Evolution of Social Norms in Common Property Resource Use. *Am. Econ. Rev.* **1996**, *86*, 766–788.
7. Osborne, M.J.; Rubinstein, A. Games with Procedurally Rational Players. *Am. Econ. Rev.* **1998**, *88*, 834–847.
8. Sethi, R. Stability of Equilibria in Games with Procedurally Rational Players. *Games Econ. Behav.* **2000**, *32*, 85–104.
9. McKelvey, R.D.; Palfrey, T.R. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38.
10. Haile, P.A.; Hortaçsu, A.; Kosenok, G. On the empirical content of quantal response equilibrium. *Am. Econ. Rev.* **2008**, *98*, 180–200.
11. Cárdenas, J.C. Norms From Outside and From Inside: An Experimental Analysis on the Governance of Local Ecosystems. *For. Policy Econ.* **2004**, *6*, 229–241.
12. Arifovic, J.; Ledyard, J. Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *J. Public Econ.* **2012**, *96*, 808–823.

13. Rodriguez-Sickert, C.; Guzmán, R.; Cárdenas, J.C. Institutions influence preferences: Evidence from a common pool resource experiment. *J. Econ. Behav. Organ.* **2008**, *67*, 215–227.
14. McKelvey, R.; Palfrey, T.R. Quantal Response Equilibria: A Brief Synopsis. In *Handbook of Experimental Economics Results*; Plott, C., Smith,V., Eds. **2008**, *1*, 541–548.
15. Brunner, C.; Camerer, C.F.; Goeree, J. Stationary Concepts for Experimental 2×2-Games: Comment. *Am. Econ. Rev.* **2011**, *101*, 1029–1040.
16. Selten, R.; Chmura, T. Stationary Concepts for Experimental 2×2-Games. *Am. Econ. Rev.* **2008**, *98*, 938–966.
17. Carpenter, J.; Cardenas, J.C. An Intercultural Examination of Cooperation in the Commons. *J. Confl. Resolut.* **2011**, *55*, 632–651.
18. Chmura, T.; Goerg, S.J.; Selten, R. Generalized Impulse Balance: An Experimental Test for a Class of 3×3 Games. *Rev. Behav. Econ.* **2014**, *1*, 27–53.
19. Stahl, D.; Wilson, P. Experimental evidence on players' models of other players. *J. Econ. Behav. Organ.* **1994**, *25*, 309–327.
20. Nagel, R. Unraveling in Guessing Games: An Experimental Study. *Am. Econ. Rev.* **1995**, *85*, 1313–1326.
21. Levine, D.K. Modelling Altruism and Spitefulness in Experiments. *Rev. Econ. Dyn.* **1998**, *1*, 593–622.
22. Fehr, E.; Schmidt, K. A Theory of Fairness, Competition and Cooperation. *Q. J. Econ.* **1999**, *114*, 817–868.
23. Bolton, G.; Ockenfels, A. A Theory of Equity, Reciprocity and Competition. *Am. Econ. Rev.* **2000**, *90*, 166–193.
24. Charness, G.; Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **2002**, *117*, 817–869.
25. Cox, J.C.; Friedman, D.; Gjerstad, S. A tractable model of reciprocity and fairness. *Games Econ. Behav.* **2007**, *59*, 17–45.
26. Rabin, M. Incorporating Fairness into Game Theory and Economics. *Am. Econ. Rev.* **1993**, *83*, 1281–1302.
27. Dufwenberg, M.; Kirchsteiger, G. A theory of sequential reciprocity. *Games Econ. Behav.* **2004**, *47*, 268–298.
28. Falk, A.; Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **2006**, *54*, 293–315.
29. Geanakoplos, J.; Pearce, D.; Stacchetti, E. Psychological Games and Sequential Rationality. *Games Econ. Behav.* **1989**, *1*, 60–79.

Representing Others in a Public Good Game

Karen Evelyn Hauge and Ole Rogeberg

Abstract: In many important public good situations the decision-making power and authority is delegated to representatives who make binding decisions on behalf of a larger group. The purpose of this study is to compare contribution decisions made by individuals with contribution decisions made by group representatives. We present the results from a laboratory experiment that compares decisions made by individuals in inter-individual public good games with decisions made by representatives on behalf of their group in inter-group public good games. Our main finding is that contribution behavior differs between individuals and group representatives, but only for women. While men's choices are equally self-interested as individuals and group representatives, women make less self-interested choices as group representatives.

Reprinted from *Games*. Cite as: Hauge, K.E.; Rogeberg, O. Representing Others in a Public Good Game. *Games* **2015**, *6*, 381–393.

1. Introduction

In many important decisions, the decision-making power and authority is delegated to representatives who make binding decisions on behalf of a larger group. In wage negotiations, for instance, representatives for the different parties meet and negotiate on behalf of their respective groups. Similarly, contracts are negotiated by representatives of each of the contract partners, international agreements on climate change or other issues are negotiated by representatives of each of the participating nations, and within firms, departments might send representatives to decide on the future strategy.

While both theoretical and empirical work within economics have focused mainly on individual decision makers and decisions made as individuals, insight from economics is commonly applied to contexts where decisions are made by individuals representing groups. If decisions made as group representatives differ systematically from decisions made as individuals, then this calls into question the external validity of research based on individual decisions to contexts where people act as group representatives. The same concern has previously been expressed by others. See for instance Cooper and Kagel [1].

Decision making by representatives is closely related to group decision making, but less studied. Group decision making refers to inter-group settings where the members of each team make decisions together. Recent research on decisions made by groups indicates that group decisions differ systematically from those of individuals.

Reviews of the literature on group decision making indicate that groups behave more in line with standard economic models than individuals: Groups are cognitively more sophisticated, have better self-control and behave in a more self-interested manner, see Charness and Sutter [2] and Kugler, *et al.* [3] for comprehensive reviews of this literature. Not all studies, however, find support for groups behaving more in line with economic models than individuals [4–13].

Studies comparing decisions made by individuals and group representatives have found that in dictator games, men (but not women) allocate less to recipients when acting as representatives compared to as individuals [14], in trust games representatives trust less and reciprocate less than individuals [15], and in the stag-hunt game representatives behave in a less risk-averse manner than individuals [16]. Song [15] argues, based on her review of this literature, that group representatives—like groups—behave in a more self-interested manner.

The above-mentioned reviews of group decisions and studies of decisions by group representatives do not include studies of behavior in public good games. The pattern from earlier studies would suggest that groups and group representatives should contribute less to public goods than individuals do in inter-individual settings. Few have studied group decisions and decisions made by group representatives in public good games[1]. Thum, Auerswald, Schmidt and Torsvik [13] study group decision making in a public good game, and find that in joint groups decisions, contributions to the public good are higher, not lower, than contributions made by individuals. In a public good game where a second person outside the lab receives the same payoff as the decision maker, however, Humphrey and Renner [20] find that contributions decrease when the second person is a friend the decision makers has brought to the lab, while contributions are not affected when the second person is an anonymous stranger. A closely related literature on delegation has examined how representatives make contribution decisions in public good games on behalf of all participating players [21–24]. In this literature, decision power is delegated to a representative (either designated [23] or elected [25]) who decides the contribution decisions of the entire group and thus determines the final outcome. The representatives are found to contribute higher amounts to the public good, giving a final outcome closer to social optimum compared to when individuals make their own contribution decisions.

Two factors that seem to matter for decisions made by group representatives, is gender and whether decisions are anonymous or made in public. As mentioned above, men and women have been found to respond differently to behaving as

---

[1] A different, but also somewhat related literature studies leading-by-example in public good games, where one person makes his contribution decision before the other group members [17–19].

representatives, compared to as individuals, in dictator games. Song, Cadsby and Morris [4] compare dictator behavior in an inter-individual dictator game and an inter-group dictator game. They found that women gave away slightly, but not significantly, more in the role as representatives compared to as individuals, while men give away significantly less as representatives. Men and women may also respond differently to having their decisions being observed by in-group and out-group members: women have been found to cooperate more when observed by their in-group, while men cooperated less [26].

In this paper, we compare contribution decisions made by individuals in inter-individual public good games with contribution decisions made by individuals on behalf of a three-person group in inter-group public good games. The study has a within-subject comparison of decisions made as individuals and as representatives. In addition, the study has a between-subject comparison of the impact of gender and anonymity. We chose the public good game for our study because important decisions in public good situations are made by group representatives, and because decisions by group representatives in public good games have not yet been studied much.

Our results show firstly that decisions made as individuals and representatives in a public good game settings differ. Secondly, we find that this difference is primarily driven by how women respond to the change in role from individual to representative for the group. Finally, we do not find support for representatives being more self-interested than individuals. To the contrary we find that men behave in an *equally* self-interested manner in the role as representatives and individuals, while women behave in a *less* self-interested manner as representatives compared to as individuals.

Our paper makes the following contributions: (1) it contributes to the literature on group decisions and decisions made by group representatives by providing a study of behavior in a public good situation; and (2) it refines the notion that group representatives are more self-interested than individuals by providing an example where this notion does not hold true.

The remainder of the paper is organized as follows: Section 2 presents the experimental design, Section 3 presents the results and Section 4 presents a concluding discussion.

2. Experimental Section

The experiment is designed to compare the choices of individuals in inter-individual public goods with the choices of the same individuals acting as group representatives who decide on behalf of their group in inter-group public good games. The experiment has a within-subject comparison of decisions made by

individuals and representatives, and a between-subject comparison of the impact of gender and anonymity.

An example can illustrate the decision environments in inter-individual public goods and inter-group public goods. In the inter-individual public good setting, three persons collaborate on a group project and must each decide how much effort they themselves will put into the project. In the inter-group public good setting, the three individuals are replaced by three groups with three persons in each group. The three groups collaborate on a common group project, and a representative from each of the three groups meet and decides how much effort his or her team will put into the group project.

Before we present the procedures of the experiment, we will present the details of these two public good settings.

2.1. Decisions as Individuals

The inter-individual public good game was a standard linear three-person public good game [27]. Participants were given an endowment of 60 Norwegian kroner (NOK) (~$10). The game was explained using the concept of a "doubling bucket". The three members of a group shared a doubling bucket, and each subject could decide how much of his or her endowment to put in the doubling bucket and how much to keep. All money placed in the bucket was doubled and divided equally between the group members, giving the monetary payoff function provided in Equation (1) below. Subjects made simultaneous-move contributions to the doubling bucket, stated as shares of the endowment in 10% increments.

$$\pi_i^I = e(1 - c_i) + \frac{2}{3}\sum_{i=1}^{3} ec_i \tag{1}$$

where c_i =is individual i's contribution as share of the endowment e.

Since contributions give a return of 2/3, the Nash equilibrium in the absence of social preferences consists of no contributions and each participant getting a payoff of e.

2.2. Decisions as Group Representatives

The inter-group public good game consists of three groups of three individuals who shared a doubling bucket. Each group member made a contribution decision on behalf of his or her three-person group. This decision was made in private, the identity of other group members was unknown, and there was no communication between group members. One of these three decisions was chosen at random and implemented as the group's contribution decision. Each group member had an endowment of 60 NOK, giving the group representative $3 \times 60 = 180$ NOK at

his or her disposal. Thus the group representative could contribute an amount between 0 and 180 to the doubling bucket on behalf of his or her group. The group representative was asked to make the following decision:

> Your task is to answer the following question: How much of your group's money do you want to put in the bucket on behalf of yourself and the two others in your group?

The contributions were stated as shares of the group's total endowment in 10% increments. By design, therefore, all three members of each group contributed the same amount.

All money placed in the bucket was doubled, and split equally between the total of nine members in the three groups sharing the bucket. The payoff function for group representative i is presented in Equation (2) below.

$$\pi_i^G = \frac{1}{3}\left[3e(1-c_g) + \frac{2}{3}\sum_{g=1}^{3} 3ec_g\right] = e(1-c_g) + \frac{2}{3}\sum_{g=1}^{3} ec_g \tag{2}$$

where $c_g \in [0,1]$ is the contribution share decision on behalf of group g stated as percentage of the group's total endowment $3e$.

The inter-group public good game is a scaled up version of the inter-individual public good game. The per-capita return to the public good is by construction identical between that of the inter-individual and the inter-group public good, so is the Nash Equilibrium in the absence of social preferences.

2.3. Experimental Procedures

The experiment consisted of four sessions and was conducted at the University of Oslo. Each session included 27 subjects, adding up to a total of 108 subjects and 324 contribution decisions. Each subject participated only once. The subjects were recruited from lectures attended by first-year students from a number of faculties of science at the University of Oslo, Norway. Fifty (46.3%) of the subjects were female. The distribution of subjects across sessions is presented in Table 1. The experiment was programmed in z-tree [28].

The experiment consisted of a training phase, a decision phase, a feedback phase and a post-experimental questionnaire. When subjects arrived at the lab, a random draw decided the seating in the lab. The *training phase* started after the general instructions[2] were read out load and before subjects started making

[2] See the on-line Supplementary file, section 3 for a translated copy of the experimental instructions and Supplementary file, section 4 for the screen-shots from the experiment.

their decisions. In the training phase subjects could test out various hypothetical contribution decisions of three fantasy players and observe how this affected payoffs.

Table 1. Descriptive statistics of subjects by session: number of subjects, female share and mean age.

Session	Number of Subjects	Female Share	Mean Age
1	27	0.56	21.9
2	27	0.56	20.4
3	27	0.41	21.7
4	27	0.33	21.4
Total	108	0.46	21.4

The *decision phase* consisted of three one-shot public good games. As we are interested in measuring the difference in contributions by individuals and group representatives, we employed a "within-subject" design, such that all subjects in the experiment made decisions both as individuals and as representatives. For half of the subjects, the first public good game was an inter-individual public good game where they made their contribution decision as an individual, the second public good game was an inter-group public good game where they made a contribution decision as a group representative, and the third and last public good game was again an inter-individual public good game where they decided as an individual. For the other half of the subjects, the order was reversed, such that the first public good game was an inter-group public good game, the second an inter-individual public good game, and the third an inter-group public good game. This crossover design with an A-B-A (individual-representative-individual) *versus* a B-A-B (representative-individual-representative) pattern makes it possible to correct for potential order effects. The experiment had a perfect stranger design [29], such that subjects were not re-matched with the same individuals twice.

After the three public good games were completed, there was a *feedback phase*. Before this, subjects received no feedback on outcomes from the games played. Postponing all feedback to after all public good games were completed ensured that outcomes from earlier games did not affect choices in later games. In the feedback phase subjects were reminded about their own individual contribution decisions and informed about their payoffs from each of the three public good games. In addition, they were informed about what their group's contribution was in inter-group public good games, and whether their decision was the one drawn to be the group representative.

The experiment had a between-subject comparison of anonymous and public decisions. Half of the subjects from each of the decision orders explained above (ABA and BAB) were in anonymous treatments and the other half in public treatments. In anonymous treatments the subjects did not know the identity nor the decisions made

by the other subjects. In public treatments the subjects knew there was a positive probability that they would have to write their group number and contribution decision on a flip-over chart at the end of the experiment in front of all the other subjects in the same session.

At the end of the feedback phase in public treatments, for each of the three public good games, three[3] of the 27 subjects who participated in the session were randomly selected and called forth to write down their contribution decision on a flip-over chart in front of the other subjects. For the inter-group public good games, the three subjects which were randomly chosen were chosen among the actual group representatives. Since "forcing" people to go public could be experienced as unpleasant and unexpected, participants in public sessions were notified of this possibility during the introduction and given the opportunity to leave with a show-up fee. No one did.

At the very end of the experiment, all subjects answered a post-experimental questionnaire covering background information such as gender, age and which faculty of science they attended at the university.

3. Results

In this section we present the behavior observed in the experiment and examine whether subjects behaved differently as individuals compared to as representatives. In particular we examine whether the difference in behavior as individuals and representatives differs between men and women.

As with any experiment with a within-subject design, there is a possibility of order effects. The design of the experiment, where choice 1 and 3 were identical treatments (see Figure 1 for the design of the experiment), allows us to test for and if necessary correct for potential order effects. A paired t-test found signs of systematic differences: the third contribution choice averaged 6 percentage points lower than the first choice, a difference that was significant at the 10% level ($p = 0.095$). For this reason, controls for the choice number (choice 1, 2 or 3) within a session are used in the regression results reported below.

Figure 2 averages an individual's contribution under identical treatments (choice 1 and 3), and illustrates average contributions by treatment and gender. This suggests an interesting pattern of gender difference: while men contribute quite similar amounts as individuals and representatives, both in anonymous and public decisions, women contribute more as representatives than they do as individuals both

[3] Only a subsample (1/9) of subjects were asked to reveal their contribution decision in front of the other subjects. This was in order to make decisions public within a reasonable time-constraint.

in anonymous and public decisions. The largest raw difference between contributions as individuals and representatives is observed in the anonymous decisions of women.

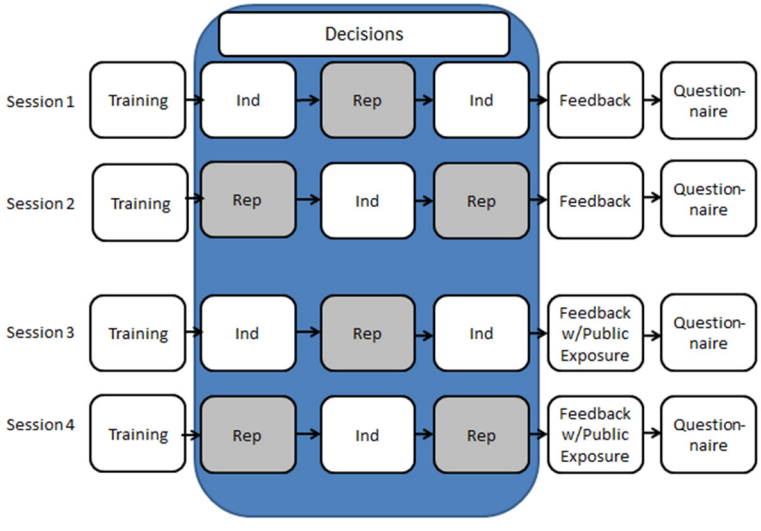


Figure 1. Outline of the experimental design. Decisions as Individuals (Ind) and Decisions as Group Representatives (Rep).

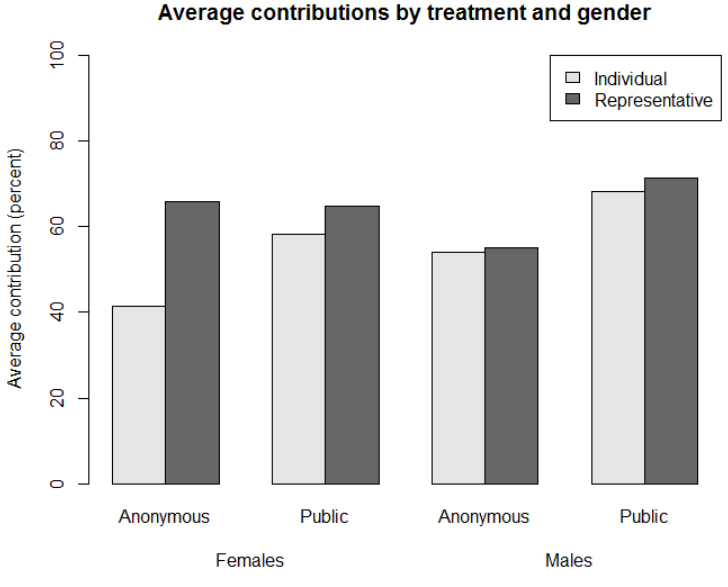


Figure 2. Average contributions (in percent of endowment) by treatment and gender.

The differences between contributions made as individuals and group representatives can be compared using non-parametric tests. Since feedback was given only after the completion of all games, a subject's choices could not be affected by the choices of the other subjects within the session. As each subject made three contribution choices, however, these choices are not independent of each other. If we nevertheless treat individual contribution decisions as independent observations, including all choices observed in one comparison, Wilcoxon rank-sum tests report a significant difference between contributions made as individuals and group representatives ($p = 0.039$). Doing the same comparison separately for men and women, Wilcoxon rank-sum tests report a significant difference between contributions made as individuals compared to group representatives for women ($p = 0.002$), but not for men ($p = 0.862$).

As in Figure 2, we can calculate the difference at the individual level between contributions made as representatives and as individuals. This gives us one (independent) observation for each participant. Comparing men and women, the Wilcoxon rank-sum test reports a significant difference between men and women ($p = 0.014$).

To further investigate the robustness of the interesting gender difference in behavior illustrated in Figure 2, Table 2 reports OLS regressions of contributions to the public good as a percentage of the endowment, with standard errors clustered at the individual level and gender specific parameters.

Gender specific parameters ease the comparison between how men and women react to behaving as group representatives *versus* as individuals. The regression equation of interest is:

$$Y_i = \sum_{s\in\{M,F\}} (\alpha^s + \beta^s_R D^s_R + \beta^s_P D^s_P + \beta^s_{R\times P} D^s_{R\times P}) + \beta_C D_C + o \tag{3}$$

here, Y_i is the public good contribution of individual i, while the D's are dummies that indicate treatment: s indicates gender, R refers to decisions made as a representative and P refers to public sessions. Interaction terms have these acronyms separated with "$\times$". In addition, there are dummies (here referred to by index C) that act as controls for the session and choice order in some of the regressions. The results are given in Table 2[4].

---

[4] See the on-line Supplementary file, section 1 for an equivalent model with gender neutral parameters.

Table 2. Regression results for outcome "Share of endowment contributed to public good". Clustering on the individual. Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

	(1)	(2)	(3)	(4)
VARIABLES	**Percent**	**Percent**	**Percent**	**Percent**
Female representative	17.30 ***	17.57 ***	22.83 ***	32.90 ***
	(5.954)	(6.046)	(5.914)	(7.916)
Male representative	2.180	1.945	6.091	9.290
	(5.230)	(5.050)	(4.848)	(7.456)
Female × Public	7.601	7.596	−1.273	8.923
	(7.252)	(7.279)	(9.566)	(11.65)
Male × Public	15.24 *	15.24 *	6.104	6.649
	(8.208)	(8.234)	(9.773)	(11.68)
Female × Public × Representative				−24.80 **
				(11.29)
Male × Public × Representative				−5.671
				(9.794)
Controls for period within session		YES	YES	YES
Controls for session			YES	YES
Female intercept	45.13 ***	45.55 ***	54.42 ***	50.63 ***
	(5.552)	(6.092)	(7.259)	(7.711)
Male intercept	53.49 ***	54.16 ***	63.60 ***	63.24 ***
	(7.171)	(7.653)	(7.979)	(8.731)
Observations	324	324	324	324
R-squared	0.741	0.745	0.757	0.761

The regressions in Table 2 include dummy variables for decisions as representatives and public decisions. The reference decision reflected in the intercept is thus an anonymous decision made as an individual in an inter-individual public good game. As can be seen from the male intercept, men contribute on average 53%–63% as individuals in anonymous decisions, depending on the model specification. Acting as representatives does not affect the contributions of men significantly in any of the specifications. When acting as individuals, the public decisions of men are 15 percentage points higher than the anonymous decisions (seen by the coefficient "Male public"), which is significant at the 10% level. When acting as representatives, the point estimates suggest that this difference between anonymous and public decisions is largely erased (see the coefficient "Male × Representative × Public").

The patterns in female decisions are strikingly different. Average female contributions in anonymous decisions as individuals lie between 45 and 51% of

the endowment depending on the model specification; which is somewhat lower than for men, but not significantly[5] so. When acting as representatives women increase their contributions substantially, and in all model specifications the increase is statistically significant at the one percent level. The estimates of the increase in female contributions as representatives from our models lie between 17 and 33 percentage points. While Model 1 only includes the treatment dimensions, Model 2 adds controls for order effects (choice 1, 2 or 3), Model 3 adds in controls for session and Model 4 allows for gender-specific interactions between the two treatment dimensions. The controls added in Models 2–4 only strengthen the estimate of increased contributions by female representatives. However, we can see from Model 4 that when decisions are revealed to others (including the members of the subgroup), women still contribute more as representatives compared to as individuals, but this difference is less than under anonymous decisions (seen in the coefficient for "Female $\times$ Representative $\times$ Public"). Further robustness checks of the gender differences are reported in the on-line Supplementary file, section 2.

Summing up, we find that men behave similarly as individuals in the inter-individual public good game and as representatives in the inter-group public good game, while we find a quite substantial and robust difference in such decisions made by women. Women contribute substantially higher levels when acting as representatives in the inter-group public good game in comparison to acting as individuals in the inter-individual public good.

4. Conclusions

In this study we have compared the behavior of individuals in inter-individual public good games with the behavior of group representatives in inter-group public good games. Our main finding is that contribution decisions made as individuals differ from those made when acting as representatives for a group, but only for women. While previous research on group decision making suggests that group decisions are more self-interested than individual decisions[6] [2,3], and while it has been suggested that this is also true for decisions made by group representatives [15], our results do not fit this pattern. To the contrary, we find that men behaved in an *equally* self-interested manner, and women behaved in a *less* self-interested manner in the role as group representatives.

A handful of other papers have also found results that do not fit the "groups and representatives of groups are more self-interested" hypothesis. Some papers have failed to find a difference between the behavior of individuals and groups in

---

[5] See the on-line Supplementary file, section 1, Table S1.

[6] Although there are several exceptions to this, as referred to in the introduction.

terms of self-interested or other-regarding behavior [6,10], some find that groups behave in a more other-regarding manner than individuals [4,5,9]. The study closest to ours is a study comparing the behavior of individuals in three-person public good games with the behavior of elected representatives in a nine-person public good game [30]. Their main finding, in line with our result for women, is that the elected representatives contribute higher amounts than individuals.

At first glance, our result appears inconsistent with the gender effects presented in Song, Cadsby and Morris [14]. As mentioned in the introduction, Song, Cadsby and Morris [14] compare decisions made as individuals and as group representatives in a dictator game. They found that women gave away slightly, but not significantly, more in the role as representatives compared to as individuals, while men gave away significantly less as representatives. This differs from our results where it is women—not men—that alter their behavior. The relative change in male and female choices when moving from the individual to the representative context, however, is in the same direction: Song, Cadsby and Morris found that men behave in a more self-interested manner as representatives, while we found that women behave in a less self-interested manner as representatives.

Although our results do not support the notion that group representatives behave in a more self-interested manner than individuals, the design of our experiment suggests that the self-interest effect might be related to group dynamics. Differences in individual and group decisions can be due to the decision maker's consideration for the final outcome, other group members or be due to group dynamics within the group. By design, there were no group dynamics in our experiment: There was no group-building exercise, no communication between the group members, and the identity (and gender) was unknown to participants. If the results of our experiment should prove representative of how representatives act in public good games, it would be interesting to repeat it with interaction between group members to see if this is what drives group decisions towards self-interest.

We found that the gender of the representative is of importance for decisions made on behalf of others in a public good game. By design, we did not give participants any information about the gender of interacting partners. The gender of others may also have an effect of decisions, according to a study that found that women—but not men—were influenced by the gender of the opposing players in a public good game [31]. However, a study comparing individual and group decision making in a trust game found no effect of gender composition on group decisions [6]. An interesting avenue for future research is whether and how the gender composition of interacting partners influences behavior.

In the introduction we expressed concern for using insight based on decisions made by individuals to contexts where decisions are made by representatives on behalf of others. Our results give reason to take this concern seriously.

Acknowledgments: This project was funded by the Research Council of Norway, grant number 193703. While carrying out this research, the authors have been associated with CREE—Oslo Centre for Research on Environmentally friendly Energy. We are grateful for valuable comments from Kjell Arne Brekke, Ragnhild Haugli Bråten, Åshild Auglænd Johnsen, Jo Thori Lind and Hans Olav Melberg on earlier drafts; and to Ragnhild Haugli Bråten and Espen Stokkereit for research assistance.

Author Contributions: Both authors contributed equally to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cooper, D.J.; Kagel, J.H. Are two heads better than one? Team *versus* individual play in signaling games. *Am. Econ. Rev.* **2005**, *95*, 477–509.
2. Charness, G.; Sutter, M. Groups make better self-interested decisions. *J. Econ. Perspect.* **2012**, *26*, 157–176.
3. Kugler, T.; Kausel, E.E.; Kocher, M.G. Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdiscip. Rev.: Cogn. Sci.* **2012**, *3*, 471–482.
4. Balafoutas, L.; Kerschbamer, R.; Kocher, M.; Sutter, M. Revealed distributional preferences: Individuals *vs.* Teams. *J. Econ. Behav. Organ.* **2014**, *108*, 319–330.
5. Cason, T.N.; Mui, V.-L. A laboratory study of group polarisation in the team dictator game. *Econ. J.* **1997**, *107*, 1465–1483.
6. Chaudhuri, A.; Paichayontvijit, T.; Shen, L. Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups. *J. Econ. Psychol.* **2013**, *34*, 181–194.
7. Cox, J.; Hayne, S. Barking up the right tree: Are small groups rational agents? *Exp. Econ.* **2006**, *9*, 209–222.
8. Gillet, J.; Schram, A.; Sonnemans, J. The tragedy of the commons revisited: The importance of group decision-making. *J. Public Econ.* **2009**, *93*, 785–797.
9. Gong, M.; Baron, J.; Kunreuther, H. Group cooperation under uncertainty. *J. Risk Uncertain.* **2009**, *39*, 251–270.
10. Kocher, M.; Sutter, M. Individual *versus* group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica* **2007**, *34*, 63–88.
11. Kocher, M.G.; Sutter, M. The decision maker matters: Individual *versus* group behaviour in experimental beauty-contest games. *Econ. J.* **2005**, *115*, 200–223.
12. Müller, W.; Tan, F. Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon. *Games Econ. Behav.* **2013**, *82*, 658–674.

13. Thum, M.; Auerswald, H.; Schmidt, C.; Torsvik, G. Teams Contribute More and Punish Less. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik—Session: Experimental Economics II, No.B09-V3. Available online: https://www.econstor.eu/dspace/bitstream/10419/100537/1/VfS_2014_pid_105.pdf (accessed on 17 September 2015).
14. Song, F.; Cadsby, B.C.; Morris, T. Other-regarding behavior and behavioral forecasts: Females *versus* males as individuals and as group representatives. *Int. J. Confl. Manag.* **2004**, *15*, 340–363.
15. Song, F. Trust and reciprocity behavior and behavioral forecasts: Individuals *versus* group-representatives. *Games Econ. Behav.* **2008**, *62*, 675–696.
16. Charness, G.; Jackson, M.O. The role of responsibility in strategic risk-taking. *J. Econ. Behav. Organ.* **2009**, *69*, 241–247.
17. Guth, W.; Levati, M.V.; Sutter, M.; van der Heijden, E. Leading by example with and without exclusion power in voluntary contribution experiments. *J. Public Econ.* **2007**, *91*, 1023–1042.
18. Levati, M.V.; Sutter, M.; van der Heijden, E. Leading by example in a public goods experiment with heterogeneity and incomplete information. *J. Confl. Resolut.* **2007**, *51*, 793–818.
19. Moxnes, E.; van der Heijden, E. The effect of leadership in a public bad experiment. *J. Confl. Resolut.* **2003**, *47*, 773–795.
20. Humphrey, S.J.; Renner, E. *The Social Costs of Responsibility*; CeDEx discussion paper series; School of Economics, University of Nottingham: Nottingham, UK, 2011.
21. Andreoni, J.; Gee, L.K. Gun for hire: Delegated enforcement and peer punishment in public goods provision. *J. Public Econ.* **2012**, *96*, 1036–1046.
22. Bernard, M.; Dreber, A.; Strimling, P.; Eriksson, K. The subgroup problem: When can binding voting on extractions from a common pool resource overcome the tragedy of the commons? *J. Econ. Behav. Organ.* **2013**, *91*, 122–130.
23. Bolle, F.; Vogel, C. Power comes with responsibility—Or does it? *Public Choice* **2011**, *148*, 459–470.
24. Hamman, J.R.; Weber, R.A.; Woon, J. An experimental investigation of electoral delegation and the provision of public goods. *Am. J. Political Sci.* **2011**, *55*, 738–752.
25. Hamman, J.R.; Loewenstein, G.; Weber, R.A. Self-interest through delegation: An additional rationale for the principal-agent relationship. *Am Econ. Rev.* **2010**, *100*, 1826–1846.
26. Charness, G.; Rustichini, A. Gender differences in cooperation with group membership. *Games Econ. Behav.* **2011**, *72*, 77–85.
27. Ledyard, J.O. Public goods: A survey of experimental research. In *The Handbook of Experimental Economics*; Kagel, J.H., Roth, A.E., Eds.; Princeton University Press: Princeton, NJ, USA, 1995; pp. 111–194.
28. Fischbacher, U. Z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.

29. Fehr, E.; Gächter, S. Fairness and retaliation: The economics of reciprocity. *J. Econ. Perspect.* **2000**, *14*, 159–181.
30. Kocher, M.G.; Tan, F.; Yu, J. *Providing Global Public Goods: Electoral Delegation and Cooperation*; Working Paper of the Max Planck Institute for Tax Law and Public Finance; Max Planck Institute for Tax Law and Public Finance: Munich, Germany, 2014.
31. Greig, F.; Bohnet, I. Exploring gendered behavior in the field with experiments: Why public goods are provided by women in a nairobi slum. *J. Econ. Behav. Organ.* **2009**, *70*, 1–9.

Section 2

The Role of the Decision-Making Regime on Cooperation in a Workgroup Social Dilemma: An Examination of Cyberloafing

Brice Corgnet, Roberto Hernán-González and Matthew W. McCarter

Abstract: A burgeoning problem facing organizations is the loss of workgroup productivity due to cyberloafing. The current paper examines how changes in the decision-making rights about what workgroup members can do on the job affect cyberloafing and subsequent work productivity. We compare two different types of decision-making regimes: autocratic decision-making and group voting. Using a laboratory experiment to simulate a data-entry organization, we find that, while autocratic decision-making and group voting regimes both curtail cyberloafing (by over 50%), it is only in group voting that there is a substantive improvement (of 38%) in a cyberloafer's subsequent work performance. Unlike autocratic decision-making, group voting leads to workgroups outperforming the control condition where cyberloafing could not be stopped. Additionally, only in the group voting regime did production levels of cyberloafers and non-loafers converge over time.

Reprinted from *Games*. Cite as: Corgnet, B.; Hernán-González, R.; McCarter, M.W. The Role of the Decision-Making Regime on Cooperation in a Workgroup Social Dilemma: An Examination of Cyberloafing. *Games* **2015**, *6*, 588–603.

1. Introduction

Many of the challenges organizations face when attempting to achieve their goals are social dilemmas: interdependent decisions where an individual attempting to satisfy his or her own interests in the short run conflicts with the collective's interests in the long run [1]. With the advent of the Internet, a social dilemma that has received increased attention in management studies is cyberloafing, where an employee "uses their company's Internet access for personal purposes (*i.e.*, web surfing and personal email use) during work hours" [2] (p. 675). In social dilemma terminology, cyberloafing is categorized as a social trap: a situation where an individual indulges in a short-term private benefit (e.g., using the Internet for personal use during work hours) while passing a long-term cost on to the collective; e.g., lost labor and reduced productivity to the workgroup or organization [3]. Surely, employees may use the Internet to cope with stress or to stimulate their creativity [4,5]. While recognizing the inevitable "grey area" between Internet use and abuse (cyberloafing) and acknowledging these caveats [6], we focus on clear cases of cyberloafing, which disrupt work [7] and are thus counterproductive [5].

The costs of cyberloafing to organizations can be substantial. Knights [8], for instance, reports $470 million in lost productivity to U.K. firms from workers taking 15 to 30 min of company time to place a bet using online gambling websites. Another study reports that workers using Facebook and Twitter are interrupted once every 10.5 min by instant messages and tweets, taking about 23 min after using a social media website to refocus their attention entirely back to their work and costing their companies about $4500 per worker every year [9]. These monetary figures leave managers with the task of increasing cooperation from cyberloafers.

Van Lange, Balliet, Parks and Van Vugt's [10] recent review on social dilemmas suggests two decision-making regimes that managers can use to encourage cooperation: autocratic decision-making and group voting. Drawing from normative decision theory [11], we seek to answer the research question: What effects do group decision-making and autocratic decision-making structures have on cyberloafing and subsequent worker performance in workgroups?

We address this question in a laboratory environment where workgroups of nine performed a data calculation and entry task while supervised by a monitor. In addition to a control condition, we considered autocratic and group voting regimes. In the autocratic decision-making condition, a monitor would unilaterally decide whether to turn off workers' access to the Internet in the middle of the experiment. In the group voting condition, a majority vote from workers was used to decide whether workers' access to the Internet should be turned off. We find that autocratic and group voting structures reduce cyberloafing by over 50%, but only group voting boosts the cyberloafer's subsequent work productivity by a substantive (and statistically different) amount: 38%. Unlike the autocratic regime, group voting leads to workgroups outperforming the control condition where cyberloafing could not be stopped.

The remainder of the paper is outlined as follows. First, we review the literature pertinent to social dilemmas, decision regimes and cyberloafing and state our hypotheses. Second, we summarize our research methodology. Third, we report the results of our hypothesis testing both in terms of statistical and practical significance. Lastly, we conclude with discussing the theoretical and practical implications of our results for social dilemma theory and organizations.

2. Theoretical Background and Hypotheses

Vroom and Yetton's [11] seminal taxonomy of organizational decision-making regimes outlines two extreme paths that an organization may take when solving a problem. At one extreme is the autocratic decision-making regime where decisions are made by a manager or boss about what those facing the decision situation can do [12]. The motivation behind an autocratic decision-making regime is efficiency. A manager, being aware of a cyberloafing problem, can take steps to avoid collective

ruin brought upon by persistent exploiters [13–15]. The effectiveness of Internet abuse detection and disciplining actions within an autocratic decision-making regime rests on general deterrence theory [16–18]. In particular, an autocratic structure allows for a manager to impose an organizational policy that limits Internet access to stop current and future cyberloafers from exploiting other group members [19].

However, the centralized decision rights of an autocratic manager may come at a cost. Most workers value autonomy in their jobs [20] and may feel spite toward those who dictate terms about what they can do [21]. In fact, workers who slack on the job may slack even more out of spite when unilateral actions are taken to force their cooperation [22]. In relation to cyberloafing, losing access to the Internet by an autocratic authority may lead the cyberloafer to not only find something else to pass the time (other than work), but exert even less effort out of spite on the job.

> *Hypothesis 1*: After the implementation of the Internet policy, the autocratic decision-making regime will fail to increase individual production by cyberloafers compared to the control condition where cyberloafing cannot be curbed. As a result, workgroup production under an autocratic decision-making structure will fail to surpass that of the control condition.

At the other end of Vroom and Yetton's [11] taxonomy is a group decision-making regime. As a form of employee empowerment [23], a group decision-making regime places decision-making rights in the hands of group members and their manager [24]. As reviewed by Van de Ven and Delbecq [25], group decision-making regimes can vary in the extent to which group members interact. At one extreme, group members may have unstructured discussions about a problem the group faces and develop solutions through conversation. At the other extreme, group members do not converse, but merely vote independently about proposed solutions to a problem. The current research focuses on this second minimal group decision process, which we refer to as a group voting regime. Group voting resembles the democratic leadership style studied by Lewin, Lippitt and White [26]. In relation to cyberloafing, a group voting regime lets workgroups decide through a vote to turn Internet access on or off.

Previous work on employee empowerment suggests that a group voting regime increases workgroup performance [27,28]. The reason for this positive relationship may be that shared decision rights among workgroup members give these workers a sense of ownership over their work and trust that each worker will do his or her share of the labor [29]. Furthermore, a group voting regime may positively affect group members who disagree with the majority about what group members can do on the job. Procedural justice research reminds us that workers often value the procedures for reaching an outcome more than the outcome's value [30]. To capitalize

on this asymmetry, the administration may give employees voice over the design of their work; e.g., voting about Internet access [31]. Worker productivity increases after merely giving employees voice about rules on their jobs [27], because choice over elements of work increases worker-perceived procedural justice and subsequent work effort [32]. Van den Bos, Vermunt and Wilke's [33] research on procedural justice suggests that a group member's commitment and motivation to work will be higher when given voice compared to when not given voice, even when the outcome of a group voting conflicts with that member's preferences or previous behavior.

> *Hypothesis 2*: After the implementation of the Internet policy, the group voting condition will lead to higher levels of individual production among cyberloafers compared to the control and autocratic decision-making conditions. As a result, workgroup production under the group voting condition will be higher than in the control and autocratic conditions.

3. Method

3.1. Sample

Participants were 220 students recruited from a subject pool of about 2000 undergraduate and graduate students at a university in the Western United States. Participants received a $7 show-up fee plus the opportunity to earn more money for participation in a 2.5-h experiment. We conducted 6 sessions in the control and group voting conditions and 10 sessions in the autocratic condition. Participants earned $32.50 on average, which includes the show-up fee. Our participants thus earned on average $13 per hour, which compares to average earnings of data-entry clerks in the United States, which were $13.37 per hour $\pm$ $2.75 [34] at the time of the study.

3.2. Design and Procedure

We employed a one-way factorial design with three conditions: a control condition, an autocratic decision-making condition and a group voting condition. Each participant was randomly assigned to one of the three conditions summarized in Table 1. In the control condition, participants could use the Internet at any time during the experiment. In the other two conditions, Internet access could be turned off after the second period. In the autocratic decision-making condition, one of the participants, the monitor, decided unilaterally whether to turn off or maintain Internet access. In the group voting condition, Internet access was turned off if the majority of workers (five out of nine) voted to do so.

3.2.1. Instruction Period

Upon arrival at the laboratory, participants were directed to private cubicles and asked to read a set of computerized instructions. Since the instructions were somewhat involved [35], participants had 20 min to read them, with a timer displayed on a large screen at the front of the laboratory. The instructions indicated that they were one of ten members of a workgroup; the workgroup would undertake a 1 h and 40-min task, broken up into five 20-min periods. Each member would work on the task, separately and in isolation, but their earnings would be calculated based on group performance.

Table 1. Summary of conditions.

Condition	Description	No. of Sessions (Participants)
Control	Internet access was maintained by the experimenter after Period 2.	6 (60)
Autocratic decision-making condition	The monitor unilaterally decided whether to turn off or maintain Internet access after Period 2.	10 (100)
Group voting condition	Workers voted on whether to turn off or maintain Internet access after Period 2. The decision selected by the majority of workers was implemented.	6 (60)

Three minutes before the end of the instruction period, the experimenter announced the time remaining and handed out a printed summary of the instructions. None of the participants asked questions or requested extra time. At the end of the instruction period, the experiment was launched from the experimenter's room.

All conditions involved the same number of participants (nine workers and one monitor), so as to be able to compare production patterns across conditions. The person who was assigned the role of monitor kept this role for the entire experiment. We conducted six sessions of 10 participants for all conditions, but the autocratic decision-making condition, for which we conducted a total of 10 sessions. More sessions were conducted for the autocratic decision-making condition so as to collect more observations on voting decisions, since, in this case, only the monitor voted on future Internet access, leaving us with only one vote per session.

3.2.2. Software

The experiment was conducted using the Virtual Organizations software developed by CYDeveloper LLC. The software facilitates a multi-party team task, controlled centrally by an experimenter.

3.2.3. The Work Task

Adapted from previous research using summation tasks [36], the work task was a particularly long and laborious task intended to resemble the monotony that can accompany organizational life and prompt Internet usage. The task required summing up tables of 36 numbers without using a pen, scratch paper or calculator (see Figure 1). Participants could sum as many tables as they wanted so that the work task was never interrupted during a period. After completion of a table, and whether the summation was correct or not, a new table appeared on the screen.

Each table had six rows and six columns of randomly-generated numbers between zero and ten. Before providing the grand total, participants had to provide a separate subtotal for all of the 12 rows and columns. Calculating these subtotals did not directly generate earnings, but could help participants compute the grand total, which generated a 40¢ profit to the group only if the grand total was correct. If the grand total was incorrect, a 20¢ penalty was deducted from group production. So that participants could not sabotage other group members' production, penalties only applied when the worker who completed the table incorrectly had produced a positive amount sufficient to bear the penalty. After completing a table, participants learned whether their answers were correct and how much money they earned. At the end of each period, participants learned the total amount of money generated by all participants' efforts on the work task. Individual earnings were calculated such that each participant obtained an equal share of 10% of workgroup production similar to a gainsharing plan. The gainsharing design feature induces interdependence among participants, as their performance on the task not only affects their individual earnings, but also the other participants' earnings. The current setting is a social dilemma, as each member of the group can increase overall performance at their own cost of effort.

	Column1	Column2	Column3	Column4	Column5	Column6	Sum Row:
	3.00	6.00	3.00	0.00	6.00	0.00	
	10.00	5.00	1.00	5.00	2.00	3.00	
	8.00	3.00	5.00	4.00	8.00	7.00	
	1.00	6.00	0.00	9.00	8.00	0.00	
	3.00	7.00	0.00	8.00	10.00	4.00	
	3.00	10.00	10.00	6.00	10.00	0.00	
Sum Column:							

Figure 1. Sample work task.

In addition to the work task earnings, participants received a fixed wage of $2.40 per period, which was obtained by clicking on a yellow box at the bottom of the participants' screens. The fixed wage was implemented to mimic real work environments in which pay for performance is only limited to a portion of the wage. In our experiments, about half of total pay was earned on the work task, while the remainder corresponded to fixed pay.

3.2.4. Internet

At any point during the experiment, all participants were told they could switch from the work task to Internet browsing. Within the bounds of university policy, participants could use the Internet however they liked, including email. Their confidentiality was assured and maintained, but the software tracked the exact amount of time spent on each activity. Although participants could not complete the work task while browsing the Internet, switching was quick and easy. Through the action menu, participants returned to either the last Internet page or the last number table that they had seen. If participants chose the Internet, the work task window was replaced by an Internet window (embedded in the software; see Figure 2). Students were not allowed to bring cell phones into the lab, so that Internet browsing, if any, was embedded into the experimental platform.

At the end of the second period, depending on the condition, Internet access was either maintained or removed. In the control condition, Internet access was maintained after the second period. In the group voting and autocratic decision-making conditions, organizational members decided on whether Internet access should be turned off after Period 2. Therefore, even if the Internet were turned off, participants could still loaf on the job by sitting idle.

Figure 2. Embedded Internet screen.

3.2.5. Monitoring

In all conditions, one of the 10 participants referred to as "*C*" (the rest of the workers were referred to as "*B*s") was given the ability to watch everyone else, and everyone else was aware of this ability. If the monitor selected the monitoring option from the action menu, he or she used a separate screen to choose whom to monitor (anywhere from one to all other participants) and to actually perform the monitoring. For each selected participant, a column in a table listed their activities (e.g., switched to Internet, provided a subtotal), their current earnings and their percentage contribution to the workgroup total (see Figure 3). For example, the first row in Figure 3 informs the monitor that Participant B13 just switched to the Internet screen and that B13 had produced 40¢ on the task thus far (that is, after 13 min and 18 s of the 20-min period). The production of Participant B13 corresponded to 33% of workgroup production (120¢) thus far. Participants who were being monitored saw the figure of an eye and a text message indicating that "*C* is watching you", independently of the activity they were undertaking. The figure and message resemble the common practice of notifying employees when they visit restricted websites. Notification systems are becoming popular, as illustrated by the team application developed by the largest online marketplace: oDesk. oDesk allows employers to overtly monitor freelancers via webcams (the analogue of our eye image), which take frequent pictures that are immediately sent back to the employer.

Participants could spend as much or as little time as they wanted on the various activities (work task, Internet and monitoring), each of which was undertaken on a separate screen. To switch activities, participants simply chose the corresponding option from a drop-down menu at the bottom-right of their screens.

3.2.6. Voting

In the autocratic decision-making condition, the monitor unilaterally decided whether to turn off or maintain Internet access after Period 2. In the group voting condition, all workers (excluding the monitor) voted on whether to turn off or maintain Internet access after Period 2. The details of the voting process were described to participants during the instruction period.

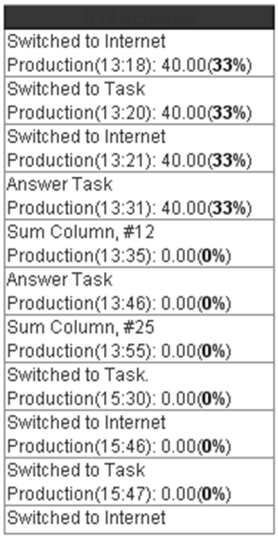


Figure 3. Monitoring screen with a zoom on Participant B13.

3.3. Measures

Individual production is defined as the total monetary amount (in U.S. Dollars) generated by a participant's answers on the work task. Workgroup production is defined as the sum of the individual production of the nine workers in the organization.

Internet usage is defined as the percentage of a participant's total time that was spent on the Internet screen. Internet usage was considered cyberloafing in this case, because time spent on the Internet was time away from the work task, costing money for participants and their colleagues. We confirm this interpretation of Internet usage in our organizational setting in which it is shown, for example, that a worker's accuracy on the summation task typically decreased after spending time browsing the Internet [37]. This would not be the case if workers used the Internet to take a break and restore concentration with the objective of increasing productivity thereafter.

4. Results

4.1. Periods 1–2: Internet Access on

We start by analyzing the first two periods of the experiment during which Internet access was available in all conditions. Average Internet usage across conditions was equal to 13.9% (SD = 0.20) of workers' available time (see Table 2). This is similar to the on-the-job Internet usage, for non-work purposes, of about

13% reported in the 2005 report by American Online and Salary.com [38]. Our cyberloafing rate is also a little less than half the rate [39] of the study of Wagner and colleagues' [40] that measures cyberloafing behaviorally in a classroom setting.

Table 2. Individual Internet usage in the first two periods across conditions (all workers, excluding the monitor).

Condition	Average Internet Usage (% of Available Time)	Proportion of Cyberloafers	Average Internet Usage for Cyberloafers
Control (n = 54)	13.8%	44.4%	29.3%
Autocratic decision-making (n = 90)	14.2%	47.8%	28.4%
Group voting (n = 54)	13.6%	48.2%	26.8%
All (n = 198)	13.9%	47.0%	28.2%

In the remainder of this section, we provide the results of t-tests in parentheses and the p-values for the corresponding non-parametric tests in brackets (either Wilcoxon rank sum or sign rank tests). The use of the Internet in the first two periods in the autocratic condition (M = 14.2%, SD = 0.20), group voting condition (M = 13.6%, SD = 0.20) and control conditions (M = 13.8%, SD = 0.20) were not statistically different from each other (all $p > 0.85$ for pairwise comparisons using t-tests, (all $p > 0.67$ for the corresponding Wilcoxon rank tests)). A large proportion of participants (53.0%) did not browse the Internet and dedicated their time to the work task. We refer to Internet users as cyberloafers and to the remaining ones as non-loafers. The proportion of cyberloafers did not vary statistically across conditions (all $p > 0.69$ for pairwise comparisons using proportion tests). Because participants faced the Internet screen at the beginning of each of the periods, those who were not categorized as cyberloafers spent some minimal amount of time on the Internet screen before switching to the task screen at the beginning of the period. Even though these participants faced the Internet screen, they did not intentionally browse the Internet. Their average Internet usage was equal to 1.3% (SD = 0.01) of their available time: about 16 s each period (no statistical differences across conditions were found; all $p > 0.82$ using t-tests). In the first two periods, workers' average production per period in the control (M = \$1.17, SD = \$1.15), autocratic decision-making (M = \$0.99, SD = \$0.89) and group voting (M = \$1.24, SD = \$1.08) conditions was not statistically different from one another (all $p > 0.14$ (all $p > 0.24$); see Table 3). Lastly, workgroup production across the control (M = \$26.26, SD = \$6.99), autocratic decision-making (M = \$25.45, SD = \$5.91) and group voting (M = \$28.19, SD = \$7.73) conditions was not statistically different from one another (all $p > 0.16$) (all $p > 0.55$).

Table 3. Average individual production per period (in U.S. Dollars) across conditions (all workers, excluding the monitor; *n* denotes the number of workers per condition).

	Periods 1 and 2			Periods 3–5		
Condition	**Cyberloafers**	**Non-loafers**	**All**	**Cyberloafers**	**Non-loafers**	**All**
Control (n = 54)	\$0.61	\$1.62	\$1.17	\$0.82	\$2.05	\$1.50
Autocratic decision-making (n = 90)	\$0.71	\$1.25	\$0.99	\$0.91	\$1.87	\$1.42
Group voting (n = 54)	\$1.02	\$1.44	\$1.24	\$1.45	\$1.73	\$1.59
All (n = 198)	\$0.77	\$1.41	\$1.11	\$1.04	\$1.89	\$1.49
Only sessions where the Internet was turned off						
Autocratic decision-making (n = 81)	\$0.70	\$1.13	\$0.92	\$0.91	\$1.78	\$1.35
Group voting (n = 45)	\$0.97	\$1.37	\$1.20	\$1.55	\$1.68	\$1.63

4.2. Voting about Internet Access

At the end of the first two periods, the workers or the monitor decided on whether to turn off Internet access in the remaining periods in the group voting and autocratic decision-making conditions. In both conditions, Internet access was shut off in all but one session, practically eradicating cyberloafing. The proportion of organizations that were able to eliminate cyberloafing in the group voting condition (5/6) did not statistically differ from the proportion of organizations achieving this outcome in the autocratic decision-making condition (9/10) (proportion test; z = 0.390, p =0.696). In the group voting condition, a majority of workers (55.6%) voted to remove Internet access.

The shutdown of Internet access in the group voting condition was not exclusively due to the non-loafers' decision to curb cyberloafing. Forty-two percent (11/26) of cyberloafers voted to turn off Internet access. Furthermore, thirty-two percent (9/28) of non-loafers voted in favor of maintaining Internet access.

4.3. Periods 3 and beyond: After Internet Access Is Voted on or off

In the last three periods, average Internet usage dropped down to 4.6% (SD = 0.11) and 5.5% (SD = 0.18) in the autocratic and group voting conditions from an initial level of 14.2% and 13.6%, respectively. The remaining presence of Internet usage is due to the fact that one session per condition maintained Internet access after Period 2. Internet usage was significantly higher in the control condition (M = 25.2%, SD = 0.30) compared to the group voting and autocratic decision-making conditions over the last three periods, and these comparisons are statistically different (all $p < 0.001$) (all $p < 0.001$). Both group voting and autocratic decision-making regimes curbed cyberloafing effectively and did not differ in Internet usage ($t(142) = 0.373$, $p = 0.710$) ($p = 0.520$).

We henceforth focus only on sessions where the Internet was turned off for a more sensitive comparison across conditions. The qualitative nature of our statistical analysis is not affected by this choice, and the analysis, which includes the two

sessions in which the Internet was maintained, is available upon request from the authors.

Individual performance increased statistically and significantly in size in the last three periods for all three conditions (all $p < 0.001$) (all $p < 0.001$). This performance increase across conditions shows the existence of learning effects (see Table 3). This finding is not surprising considering previous work reports similar learning effects in mathematical tasks [41]. The evolution of period production did vary across conditions in the case of cyberloafers, however (see Figure 4). In the case of cyberloafers, the increase in individual production between the first two periods (in which Internet access was available) and the last three periods (in which Internet access was removed) in the group voting condition (from \$0.97–\$1.55, a 59.8% increase) was near double that of the autocratic decision-making (from \$0.70–\$0.91, a 30.0% increase) and control (from \$0.61–\$0.82, a 34.4% increase) conditions. Production of cyberloafers in the last three periods of the group voting condition was statistically higher than in the first two periods ($t(18) = 3.50$, $p = 0.003$) ($p = 0.001$), whereas the difference was statistically marginal for the other two conditions ($t(23) = 1.65$, $p = 0.113$ ($p = 0.083$) and $t(39) = 1.88$, $p = 0.068$ ($p = 0.354$) for the control and autocratic decision-making conditions). Furthermore, cyberloafers' production was statistically higher in the group voting condition than in the autocratic decision-making and control conditions ($t(57) = 2.54$, $p = 0.014$ ($p = 0.017$) and $t(41) = 2.54$, $p = 0.015$ ($p = 0.021$)). There was no statistical difference between cyberloafers' performance in the autocratic decision-making and control conditions ($t(62) = 0.38$, $p = 0.700$) ($p = 0.770$). These findings are in line with our Hypothesis 1 regarding the absence of differences in individual production between the autocratic decision-making and control conditions.

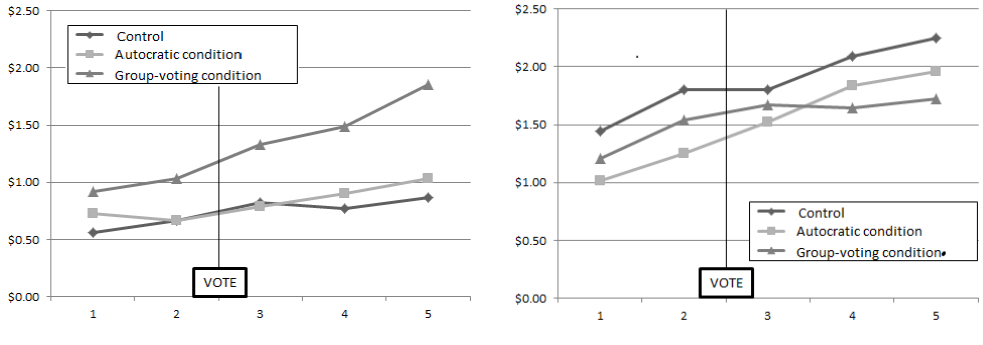


Figure 4. Evolution of the individual production of cyberloafers (left panel) and non-loafers (right panel) across periods and conditions.

Our findings also provide support to our Hypothesis 2. Non-loafers' period production increased more between the first two periods and the last three periods

in the autocratic decision-making condition (from \$1.13–\$1.77, a 56.6% increase) than in the control (from \$1.62–\$2.05, a 26.5% increase) and group voting (from \$1.37–\$1.68, a 22.6% increase) conditions. Production per period was statistically higher in the last three periods than in the first two in all three conditions ($t(40) = 6.57$, $p < 0.001$ ($p < 0.001$); $t(25) = 2.32$, $p = 0.029$ ($p = 0.013$) and $t(29) = 4.46$, $p < 0.001$ ($p < 0.001$)) for the autocratic decision-making, group voting and control conditions). Non-loafers' production thus increased in all conditions, but this increase was more pronounced in the autocratic decision-making and control conditions, where shirking by cyberloafers was also more pervasive than in the group voting condition. This observation is actually consistent with the findings of van Dijk, Sonnemans and van Winden [42], who showed that workgroups, in which social loafing was pervasive, performed surprisingly well because of the extra effort of non-loafers that aimed at compensating the particularly low level of effort of social loafers. Importantly, non-loafers' production levels in the last three periods did not statistically differ across treatments (all $p > 0.290$) (all $p > 0.380$), confirming that Internet restriction policies primarily impact those who are regular Internet users on the job; *i.e.*, cyberloafers.

In the group voting condition, workers who voted in favor of shutting down Internet access increased their individual level of production from the first two periods to the last three periods by 24.6% ($t(26) = 2.76$, $p = 0.011$) ($p = 0.002$) compared to an increase of 55.6% for those workers who did not ($t(17) = 2.97$, $p = 0.009$) ($p = 0.008$). The increase in production among workers who voted against turning off the Internet was more pronounced for cyberloafers, whose production increased by 118.4% ($t(9) = 3.35$, $p = 0.009$) ($p = 0.006$), compared to non-loafers, whose production increased by only 10.9% ($t(8) = 2.25$, $p = 0.055$) ($p = 0.475$). That is, in the group voting condition, group members did not react negatively to the implementation of an Internet policy that they opposed. Instead, they increased their production levels under the newly-implemented policy. The absence of a negative reaction is consistent with procedural justice research emphasizing the positive effect of giving employees a voice in decision-making, even when the final outcome conflicts with one's preferences [33].

Average workgroup production in the autocratic decision-making condition (M = \$42.47, SD = \$21.77) was 4.5% higher than in the control condition (M = \$40.63, SD = \$14.54). However, workgroup production in the group voting condition (M = \$51.84, SD = \$26.36) was 27.6% higher than in the control condition and 22.1% higher than in the autocratic decision-making condition. Even though these differences in production are not statistically different (all $p > 0.390$) (all $p > 0.280$), the effect for the comparison between group voting and autocratic decision-making conditions is of a moderate size (Cohen's $d = 0.40$). Our findings are consistent with our Hypothesis 1 according to which the autocratic structure will fail to increase

workgroup production statistically despite curbing cyberloafing. However, we lack support for the second part of Hypothesis 2, as the performance of workgroups using group voting structures does not statistically differ from the control condition, despite exhibiting a significant increase in production. The null statistical effect is mostly explainable by the fact that, after the second period, non-loafers' performance did not increase as much in the group voting condition compared to the control and the autocratic decision-making conditions. As mentioned above, the group voting condition is the one in which the increase in production levels was the highest for cyberloafers, while being the lowest for non-loafers. It follows that the gap in production levels between non-loafers and cyberloafers decreased significantly over time in the group voting condition (see Figure 5).

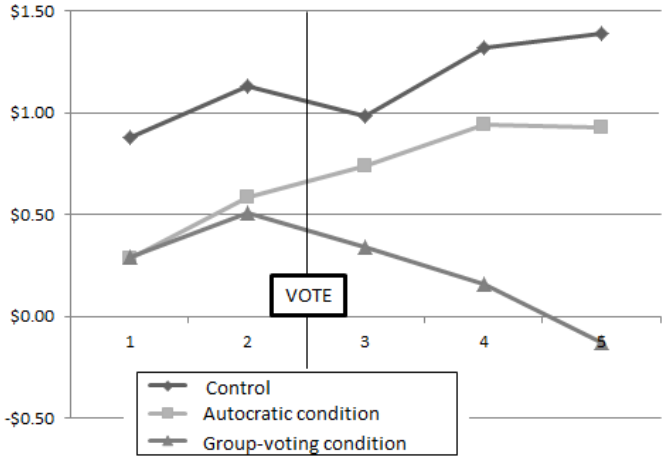


Figure 5. Difference in individual production between non-loafers and cyberloafers across periods and conditions.

By contrast, the gap in production levels increased in the autocratic and control conditions. As shown in Figure 5, the performance of cyberloafers and non-loafers did not statistically differ in the last three periods of the experiment in the group voting condition ($t(43) = 0.391$, $p = 0.698$), while cyberloafers underperformed non-loafers in the other two conditions (all $p < 0.001$) (all $p < 0.001$). The group voting condition is the only one under which the production levels of cyberloafers and non-loafers converged over time. The observed convergence in production among cyberloafers and non-loafers is especially striking given that a large proportion of workers (40.0% in the five sessions in which Internet was turned off) voted against turning off the Internet.

Without a way of making decisions about what workers could do on the job, our workgroup's performance suffered: Internet usage in the control condition increased

from 13.7% in the first part of the experiment to 25.2% in the second part of the experiment. Both autocratic and group voting structures were similarly effective in curbing cyberloafing: from an initial level of 14.2% and 13.6%, average Internet usage declined to 4.6% and 5.5% in the autocratic and group voting conditions. This finding echoes the study of Barker [43] that showed how self-managing teams ultimately imposed even more control on workers' behavior than a hierarchical system.

However, this is not all these decision-making regimes did. Organizations in the autocratic structure condition produced $56.06 on average, which is $5.67 less than organizations in the control condition, whereas organizations in the group voting condition increased their production with respect to those in the control condition up to $65.30. This difference was because of the change in behavior of the cyberloafers between the autocratic and group voting conditions.

5. General Discussion

A recent squib in *The Globe and Mail* suggests that answering "how can I stop people loafing in my workgroups?" keeps leaders awake at night [44]. The current research answers this question through comparing different decision-making regimes in a workgroup social dilemma. We found that the decision-making structure impacts not only whether people can cyberloaf, but also what the cyberloafer does thereafter. Autocratic decision-making and group voting regimes both removed the Internet form of loafing in workgroups (87.5% of the time), but only group voting let leaders have their cake and eat it too. Group voting increased a cyberloafer's work performance even when the cyberloafer had voted against turning off the Internet. An autocratic decision-making regime, by comparison, failed to increase cyberloafers' production more than the control condition after the intervention had been taken.

Our findings suggest that, despite the overwhelming monetary costs of cyberloafing identified in recent reports [8,45], managers should refrain from following an autocratic decision-making regime to restrict Internet access. Instead, organizations may consider letting employees decide on the Internet restriction policy. This may not only curb cyberloafing, but increase work performance over time.

While voting on Internet access is unlikely to lead to consensus, as it was in our study, workers who voted against the implemented measure did not react negatively. Cyberloafers increased their work performance after having Internet access restricted. In the last hour of our experiment, after the Internet restriction policy was implemented through group voting, cyberloafers produced 71.4% more than in the control condition, which is equivalent to a $1.80 increase in hourly production per clerical worker.

In conclusion, cyberloafing is real in workgroups. Cyberloafing is costly to organizations. However, cyberloafing is navigable through altering the decision-making regime in the workgroup. The key is for organizations to select a

decision-making regime that stops workers from cyberloafing while also encouraging them to work harder. The group voting structure seems to be one way for managers to address cyberloafing and also to get better sleep at night.

Acknowledgments: We are indebted to Special Issue editor Ananish Chaudhuri and two anonymous referees during the review process. We are also indebted to the Economic Science Institute for their use of their facilities during data collection and to Mark Lengnick-Hall and Dina Krasikova for their comments on previous drafts of this manuscript. The second author acknowledges financial support from the Spanish Ministry of Education [Grant 2012/00103/001], Spanish Plan Nacional I+D MCI [ECO2013-44879-R], 2014-17, and Proyectos de Excelencia de la Junta Andalucía [P12.SEJ.1436], 2014-18. This research started while the second author was visiting the Economic Science Institute at Chapman University and working at the Universidad de Granada. Portions of this paper were completed while the third author was funded by the Wang-Fradkin Assistant Professorship while at Chapman University.

Author Contributions: Matthew McCarter developed the research question; Brice Corgnet and Roberto Hernán-González conceived and designed the experiments; Brice Corgnet, Roberto Hernán-González, and Matthew McCarter wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dawes, R.M. Social dilemmas. *Annu. Rev. Psychol.* **1980**, *31*, 169–193.
2. Lim, V.K.G. The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. *J. Organ. Behav.* **2002**, *23*, 675–694.
3. Messick, D.M.; Brewer, M. Solving social dilemmas. In *Review of Personality and Social Psychology*; Wheeler, L., Shaver, P., Eds.; Sage Publications: Beverly Hills, CA, USA, 1983; Volume 4, pp. 11–44.
4. Coker, B.L. Workplace Internet leisure browsing. *Hum. Perform.* **2013**, *26*, 114–125.
5. Henle, C.A.; Blanchard, A.L. The interaction of work stressors and organizational sanctions on cyberloafing. *J. Managerial Iss.* **2008**, *20*, 383–400.
6. Lim, V.K.; Teo, T.S. Prevalence, perceived seriousness, justification and regulation of cyberloafing in Singapore: An exploratory study. *Inf. Manag.* **2005**, *42*, 1081–1093.
7. Jett, Q.R.; George, J.M. Work interrupted: A closer look at the role of interruptions in organizational life. *Acad. Manag. Rev.* **2003**, *28*, 494–507.
8. Knights, M. Gambling at work costs employers UK plc £306m. Computerworld UK, 15 January 2007. Available online: http://www.cio.co.uk/news/strategy/gambling-at-work-costs-uk-plc-306m/ (accessed on 2 November 2015).
9. Graves, J.A. The top cyberloafing activities of a distracted office worker. U.S. News and World Report, 21 March 2013. Available online: http://money.usnews.com/money/careers/articles/2013/03/21/the-top-cyberloafing-activities-of-a-distracted-office-worker (accessed on 2 November 2015).
10. Van Lange, P.A.M.; Balliet, D.; Parks, C.D.; van Vugt, M. *Social Dilemmas: The Psychology of Cooperation*; Oxford University Press: Oxford, UK, 2014.
11. Vroom, V.H.; Yetton, P.W. *Leadership and Decision-making*; University of Pittsburgh Press: Pittsburgh, PA, USA, 1973.

12. Van Vugt, M.; Jepson, S.F.; Hart, C.M.; de Cremer, D. Autocratic leadership in social dilemmas: A threat to group stability. *J. Exp. Soc. Psychol.* **2004**, *40*, 1–13.
13. Chen, X.-P.; Bachrach, D.G. Tolerance of free-riding: The effects of defection size, defection pattern, and social orientation in a repeated public goods dilemma. *Organ. Behav. Hum. Decis. Processes* **2003**, *90*, 139–147.
14. Hardin, G. The tragedy of the commons. *Science* **1968**, *162*, 1243–1248.
15. Hardin, G. Living on a lifeboat. *BioScience* **1974**, *24*, 561–568.
16. Straub, D.W. Effective IS security: An empirical study. *Inf. Syst. Res.* **1990**, *1*, 255–276.
17. Straub, D.W.; Nance, W.D. Discovering and disciplining computer abuse in organizations: A field study. *MIS Q.* **1990**, *14*, 45–60.
18. Straub, D.W.; Welke, R.J. Coping with systems risk: Security planning models for management decision making. *MIS Q.* **1998**, *22*, 441–469.
19. Ugrin, J.C.; Pearson, J.M. The effects of sanctions and stigmas on cyberloafing. *Comput. Hum. Behav.* **2013**, *29*, 812–820.
20. Laschinger, H.K.S.; Finegan, J.; Shamian, J. The impact of workplace empowerment, organizational trust on staff nurses' work satisfaction and organizational commitment. *Health Care Manag. Rev.* **2001**, *26*, 7–23.
21. Lawrence, T.B.; Robinson, S.L. Ain't misbehavin: Workplace deviance as organizational resistance. *J. Manag.* **2007**, *33*, 378–394.
22. Tripp, T.M.; Bies, R.J. *Getting Even: The Truth about Workplace Revenge—And How to Stop It*; Wiley: New York, NY, USA, 2009.
23. Martin, S.; Liao, H.; Campbell, E.M. Directive *versus* empowering leadership: A field experiment comparing the impact on task proficiency and proactivity. *Acad. Manag. J.* **2012**, *56*, 1372–1395.
24. Somech, A. The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. *J. Manag.* **2006**, *32*, 132–157.
25. Van de Ven, A.H.; Delbecq, A.L. The effectiveness of nominal, Delphi, and interacting group decision making processes. *Acad. Manag. J.* **1974**, *17*, 605–621.
26. Lewin, K.; Lippitt, R.; White, R.K. Patterns of aggressive behavior in experimentally created "social climates". *J. Soc. Psychol.* **1939**, *10*, 271–299.
27. Lawler, E.E.; Hackman, J.R. Impact of employee participation in the development of pay incentive plans: A field experiment. *J. Appl. Psychol.* **1969**, *53*, 467–471.
28. Srivastava, A.; Bartol, K.M.; Locke, E.A. Empowering leadership in management teams: Effects on knowledge sharing, efficacy, and performance. *Acad. Manag. J.* **2006**, *49*, 1239–1251.
29. Huang, X.; Iun, J.; Liu, A.; Gong, Y. Does participative leadership enhance work performance by inducing empowerment or trust? The differential effects on managerial and non-managerial subordinates. *J. Organ. Behav.* **2010**, *31*, 122–143.
30. Folger, R.; Konovsky, M.A. Effects of procedural and distributive justice on reactions to pay raise decisions. *Acad. Manag. J.* **1989**, *32*, 115–130.
31. Lind, E.A.; Tyler, T.R. *The Social Psychology of Procedural Justice*; Plenum: New York, NY, USA, 1988.

32. Caza, A.; McCarter, M.W.; Northcraft, G.B. Performance benefits of reward choice: A procedural justice perspective. *Hum. Resour. Manag. J.* **2015**, *25*, 184–199.
33. Van den Bos, K.; Vermunt, R.; Wilke, H.A.M. Procedural and distributive justice: What is fair depends more on what comes first than on what comes next. *J. Personal. Soc. Psychol.* **1997**, *72*, 95–104.
34. Salary Wizard. Available online: http://swz.salary.com/SalaryWizard/Data-Entry-Clerk-I-Job-Description.aspx. (accessed on 2 November 2015).
35. Decision-making and Cooperation in a Workgroup Social Dilemmas. 2015. Available online: http://www.sites.google.com/site/voInternetpolicy (accessed on 2 November 2015).
36. Eriksson, T.; Poulsen, A.; Villeval, M.C. Feedback and incentives: Experimental evidence. *Labour Econ.* **2009**, *16*, 679–688.
37. Corgnet, B.; Hernán-González, R.; Schniter, E. Why real leisure really matters: Incentive effects on real effort in the laboratory. *Exp. Econ.* **2014**, *18*, 1–18.
38. Malachowski, D.; Simonini, J. Wasted time at work costing companies billions. San Francisco Chronicle, 11 July 2005. Available online: http://www.salary.com/wasted-time-at-work-still-costing-companies-billions-in-2006/ (accessed on 2 November 2015).
39. World's Shortest Work Weeks 2013. Available online: http://www.money.cnn.com (accessed on 2 November 2015).
40. Wagner, D.T.; Barnes, C.M.; Lim, V.K.G.; Ferris, D.L. Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi-experiment. *J. Appl. Psychol.* **2012**, *97*, 1068–1076.
41. Charness, N.; Campbell, J.I. Acquiring skill at mental calculation in adulthood: A task decomposition. *J. Exp. Psychol. Gen.* **1988**, *117*, 115–129.
42. Van Dijk, F.; Sonnemans, J.; van Winden, F. Incentive systems in a real effort experiment. *Eur. Econ. Rev.* **2001**, *45*, 187–214.
43. Barker, J.R. Tightening the iron cage: Concertive control in self-managing teams. *Adm. Sci. Q.* **1993**, *38*, 408–437.
44. Schachter, H. Eight questions that cause leaders to lose sleep. The Globe and Mail, 5 January 2014. Available online: http://www.theglobeandmail.com/report-on-business/careers/management/eight-questions-that-cause-leaders-to-lose-sleep/article16186886/ (accessed on 2 November 2015).
45. Muller, J. Cyberloafing becoming all too common at office. Deseret News, 22 November 1998. Available online: http://www.deseretnews.com/article/664207/Cyberloafing-becoming-all-too-common-at-office.html?pg=all (accessed on 2 November 2015).

Pledges of Commitment and Cooperation in Partnerships

Lachlan Deer and Ralph-C. Bayer

Abstract: We use experimental methods to investigate whether pledges of commitment can improve cooperation in endogenously-formed partnerships facing a social dilemma. Treatments vary in terms of the individual's: (1) opportunity to commit to their partner; (2) the cost of dissolving committed partnerships; and (3) the distribution of these dissolution costs between partners. Our findings show that pledges of commitment alone can increase cooperation and welfare in committed partnerships. The introduction of relatively large and equally split costs yields similar gains. In contrast, when costs to dissolve committed partnerships fall solely on the individual choosing to break up, pledges of commitment fail to improve cooperation and welfare.

Reprinted from *Games*. Cite as: Deer, L.; Bayer, R.-C. Pledges of Commitment and Cooperation in Partnerships. *Games* **2016**, *7*, 4.

1. Introduction

We conduct an experimental study to investigate how pledges of commitment influence cooperation in a social dilemma with endogenously-formed partnerships. The social dilemma we study is a modification of the standard linear voluntary contribution mechanism (VCM) ([1]).[1] The VCM provides a simple framework that allows for a straightforward measurement of self-interested *versus* partnership-orientated behaviour. We augment the standard design by periodically introducing commitment and regrouping phases.[2] In commitment phases, subjects in uncommitted partnerships simultaneously choose whether to pledge commitment to their partner. When both partners choose to commit, a committed partnership forms. Both subjects remain in a committed partnership until the end of the game or until one (or both) chooses to dissolve the partnership during a regrouping phase. In regrouping phases, subjects choose whether to stay

[1] The review articles [2,3] provides excellent surveys of the existing literature on public goods game experiments and various modifications that aid or hinder sustaining cooperation in this environment.

[2] Previous studies of endogenous group formation [4–8] find that the introduction of endogenous group formation increases cooperation and welfare when compared to exogenous regrouping protocols (see [9] for an exception to this result). Our focus is different: we ask whether the addition of pledges of commitment in an endogenous group formation setting can provide additional increases in cooperation and welfare.

in their current partnership or to dissolve the partnership and be rematched to a new partner. Subjects whose partnership has been dissolved are rematched to new partners according to their preferences using the stable marriage mechanism ([8]).

Our experimental design allows us to disentangle the effects of the two interrelated parts a pledge of commitment entails: the pledge itself and the costs of dissolving a committed partnership. A pledge of commitment by itself could serve to increase cooperation through signalling cooperative intentions or through serving a reassuring role. However, the formation of a committed partnership rarely involves only costless, mutual pledges of commitment. Costs are typically incurred when the partnership fails and needs to be dissolved. When dissolution costs are present, pledging commitment indicates a willingness to bear a cost if the partnership is unsuccessful. This potentially strengthens the credibility of a pledge of commitment, leading to higher cooperation levels.

We also want to understand the role of the size and distribution of dissolution costs that are incurred by committed partners. Holding the distribution of costs between partners fixed, the effect of an increase in the total cost of dissolution on cooperation is ambiguous. A partner may increase contributions to avoid being broken up with and the resulting higher costs. On the other hand, choosing to dissolve a partnership is more costly, decreasing the probability a subject dissolves the partnership, providing their partner an incentive to cooperate less. The distribution of costs between partners upon dissolution should also play a role in determining cooperation levels among committed partners. When the cost burden falls asymmetrically on the instigator of the dissolution, he may be less likely to dissolve, perceiving costs of dissolution too high or the distribution unfair. Anticipating this reluctance, there is a scope for partners to behave opportunistically by decreasing cooperation.

Qualitative aspects of the experimental design seem well suited to stylized features of the marriage market and some business partnerships. In the marriage market, individuals learn about the cooperativeness of their partner in an initial dating stage and may decide to pledge commitment through marriage. Dissolving a marriage through divorce is often costly, and the size and distribution of these costs may fall differently in each relationship. Our set-up speaks to how and whether these costs may influence the ultimate success of the relationship. Business partnerships can also be thought of as an in-kind social dilemma with costly dissolution, where partners trade-off work to enhance the value of the company versus working for oneself or shirking. Dissolving these partnerships may either be cost-free via informal agreements or costly via lost investments and/or costs to retrieve these investments. These costs may fall unevenly if one partner is more invested than the other or equally if investment levels are similar for each partner.

Our results show that pledges of commitment by themselves are associated with increases in cooperation in committed partnerships and translate into improvements in welfare. The increase in cooperation comes via an ability for partners to signal cooperative intentions. Once committed, subjects act on these intentions with high contributions. When dissolution is costly, we find that committed partnerships featured higher contributions when the breakup cost was equally shared. The introduction of these costs meant that fewer subjects were willing to commit, limiting the gains in welfare. Welfare gains are only found when the total dissolution burden is relatively high. Few committed partnerships increased contributions in response to the threat of costly breakups, bringing welfare to a level equal to when committed partnerships were costless to dissolve. In contrast, committed partnerships are less cooperative than partnerships in a "commitment-free" benchmark when costs to dissolve the partnership fall asymmetrically on the instigator of the breakup. Subjects use the reluctance of their partner to dissolve to decrease contributions and free ride, leaving committed partnerships caught in a low contribution trap.

The remainder of the paper is organised as follows: Section 2 presents the research questions our experiment is designed to address. Section 3 details the experimental design. Results are contained in Section 4. Section 5 contains final remarks and concludes.

2. Research Questions

The objectives of the study are two-fold. The first objective is general: to investigate whether pledges of commitment can increase and sustain cooperation when partnership formation is endogenous. We are interested in whether a pledge itself is enough to increase cooperation and whether the presence of costs to dissolve the partnership leads to further benefits.

Research Question 1. Do pledges of commitment increase cooperation in endogenously-formed partnerships?

Research Question 2. Does the presence of dissolution costs further increase the cooperation of committed partnerships?

The second, more focussed objective is to investigate how cooperation is influenced by the size and distribution of costs attached to dissolving committed partnerships. Along the cost dimension, we vary the size and distribution of the dissolution costs between partners. We are interested in how changes in the total cost burden, holding distribution fixed, influence the cooperation of committed partnerships. We also study the effect of asymmetries in the distribution for a fixed cost burden on cooperative behaviour.

Research Question 3. How important is the total size of the cost to dissolve a committed partnership in influencing cooperation?

Research Question 4. Do asymmetric dissolution costs effect cooperation in committed partnerships?

3. Experiment Design

Each session features 32 periods of interaction, divided into eight sequences of four periods. In each sequence, all partnerships play four identical linear VCM games.[3] At the conclusion of a sequence, subjects decide to dissolve their current partnership or remain together for the subsequent sequence. Subjects whose partnership is dissolved face a preference input and regrouping phase. Those in partnerships that did not dissolve enter directly into the next sequence. In experimental treatments, subjects in new or existing uncommitted partnerships face a commitment decision at the beginning of a sequence, prior to the VCM stages. In one control treatment, we remove the commitment phase at the beginning of the sequence. An additional control treatment also removes the commitment phase and replaces the endogenous regrouping protocol with random matching. Figure 1 graphically illustrates the order of play in a typical sequence.

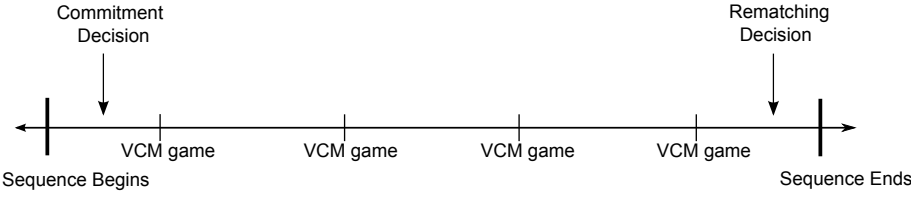


Figure 1. Timeline of a typical sequence in the game.

We now explain the three decision stages subjects face in each sequence, starting with the VCM.

3.1. The Voluntary Contribution Mechanism

The parametric structure of the VCM draws heavily on [8] and [6]. At the beginning of each period, subjects are endowed with 100 units of experimental

[3] Our experiment uses a public goods game with two-person groups rather than the more common four-player groups. This design choice is to simplify the interpretation of pledges of commitment and dissolution. In our two-player set-up, pledges of commitment can be interpreted as a pledge to one person and dissolution decisions as the result of the behaviour of one person: the subject's partner. This simplification is not possible in larger groups, where decisions are made in response to group behaviour, and is not solely attributable to the decisions made by one group member. We find similar improvements in welfare between exogenous and endogenous group formation protocols in our partner treatments compared to the existing literature using four-player or more groups, indicating that effects on cooperation in our treatments may be generalizable to larger groups.

currency. A subject chooses how to allocate the endowment between a private account, which is his or hers to keep, and a partnership account, which yields profit for both partners. We refer to the amount allocated to the partnership account as the subject's contribution. Each unit contributed to the group account, $c_i \in [0, 100]$, generates 0.8 units of profit for each partner. The profit arising to each individual from the partnership's contributions is $0.8(c_i + c_j)$. Payoffs are symmetric across partners, and allocation decisions are made simultaneously. The subject's profit from each VCM stage is:

$$\Pi_i(c_i, c_j) := 100 - c_i + 0.8(c_i + c_j),\ i, j \in \{1, 2\}, i \neq j$$

After the contribution decisions have been made, they are revealed to both partners, and the game continues. Each partner is informed about the total contribution to the partnership account, the individual contributions of both partners and their own profit from that period.

The payoff structure generates a social dilemma, because the net return from contributing one unit into the partnership's account is negative (-0.2), whilst the net social return of one unit contributed is positive (0.6). Average contributions are an indicator for a partnership's success at overcoming the social dilemma because aggregate partnership welfare (W) increases linearly with the average contribution of each partner,

$$\begin{aligned} W &= 200 + (0.6)(c_i + c_j) \\ &= 200 + 1.2\bar{c} \end{aligned}$$

3.2. Pledges of Commitment

Subjects in experimental treatments face a commitment decision at the beginning of each sequence of four VCM games. Each partner in a newly-formed partnership or surviving uncommitted partnership has the opportunity to pledge commitment to their current partner. Subjects are asked simultaneously whether they want to pledge commitment to their partner. If both partners choose to pledge commitment, a committed partnership is formed. Partners were informed of the outcome and remained in a committed partnership until it was dissolved in a regrouping phase or the experiment ended. In all other situations, subjects remained in an uncommitted partnership. When one partner wanted to commit and the other did not, the willing partner was informed that his or her counterpart was "not willing to commit right now." If both partners decided not to commit, both were informed of the outcome. Subjects then progressed to the next four VCM periods.

3.3. Regrouping

At the conclusion of each sequence, all subjects could be regrouped. In one control treatment, all subjects were regrouped randomly by the computer and continued onto the next sequence. In all other treatments, subjects could choose to dissolve their current partnership and be regrouped.[4] A partnership was dissolved if one or both partners chose to dissolve. Subject's whose partnership dissolved enter into a pool of "singles" to be regrouped. There was no cost of dissolving an uncommitted partnership. The cost of dissolving committed partnerships varied across treatments, as shown in Table 1.

All "single" subjects were regrouped according to the stable marriage algorithm [10]. Subjects were shown information about the average contributions from the previous sequence of four VCM games of "single" subjects to form preferences over potential partners.[5] Subjects ranked potential partners in order of desirability.[6] When the ranking process was complete, the matching algorithm used these preferences to produce a stable matching. Subjects were informed that they had been regrouped and entered into the next sequence without information about with whom they had been matched. They could only use indirect inference by observing their partner's contributions over the next sequence.

The mechanism underlying group formation is quite complex. For this reason, instructions provided to subjects do not fully explain how the algorithm works to avoid confusion and noisy regrouping behaviour. The regrouping procedure is described in the instructions as follows:[7][8]

"The computer will collect the rankings from every 'single' and rematch all 'singles' according to these rankings. The partner you are matched with is

---

[4] In Period 1 of each session, "types" and an initial grouping were randomly determined by the computer. Each initial grouping features two partners, each of a different type. In a regrouping phase, only subjects with different types could form a partnership. Subjects could only see potential partners who were of a different type from themselves.

[5] The decision to provide information on the average contributions of subjects over the previous sequence reflects an important compromise. When only contribution history from the previous sequence is used as information in regrouping phases, subjects are able to escape any reputation attained during sequences that occurred earlier. The advantage of the approach implemented is that both accidents and interactions with low contributing partners are removed from a subject's history quickly.

[6] Subjects rank potential partners by typing a number into a box next to the information about each of them. This preference-based regrouping follows [8] and [7].

[7] See [11] for a discussion of the performance stable marriage mechanism in a linear VCM with endogenous regrouping. He explains the workings of the mechanism to the subjects using almost identical language and finds that subjects generally base their preferences on contribution levels, ranking potential partners based on their contribution history.

[8] Sample instructions from the experiment can be found in the Online Appendix.

determined by your preferences and the preferences of all other 'singles.' The computer is programmed to give you the best partner available."

Table 1. Summary of treatments implemented.

Treatment	No. of Sessions	No. of Subjects	Endogenous Regrouping	Option to Commit	Cost to Dissolve Committed Partnership	Share of Cost Paid by Each Partner: Unilateral Dissolution	Share of Cost Paid by Each Partner: Bilateral Dissolution
Random Matching	2	30	No	No	-	-	-
No Commitment	3	56	Yes	No	-	-	-
Costless Pledge	3	58	Yes	Yes	0	-	-
Asymmetric Cost	3	60	Yes	Yes	400	Instigator pays \$400 Other pays \$0	Both pay \$200
Shared Low Cost	3	50	Yes	Yes	400	Both pay \$200	Both pay \$200
Shared High Cost	3	60	Yes	Yes	800	Both pay \$400	Both pay \$400

3.4. Treatments

We implement six treatments to investigate the research questions posed in Section 2. Treatment differences are summarized in Table 1. The Random Matching treatment and No Commitment treatment serve as controls. Four experimental treatments add periodic commitment opportunities to the No Commitment treatment and vary in the cost structure of dissolving committed partnerships.

Differences in contribution behaviour between No Commitment and Costless Pledge identify how a pure pledge of commitment influences cooperation in endogenously-formed partnerships (Research Question 1). Remaining differences between Costless Pledge and other experimental treatments highlight the importance of dissolution costs in enhancing cooperation (Research Question 2). Comparing cooperation levels between Shared Low Cost and Shared High Cost isolates the role of the total cost burden (Research Question 3). Finally, the role of asymmetry in costs is seen by comparing Shared Low Cost and Asymmetric Cost (Research Question 4).

3.5. Implementation

We conducted seventeen sessions, two sessions of the Random Matching treatment and three sessions for all other treatments described in Table 1. Between 16 and 20 subjects participated in each session, for a total of 314 participants. All sessions were conducted at the Adelaide Laboratory for Experimental Economics (AdLab) at the University of Adelaide. The experiment was computerized, and scripts were programmed using the z-Tree platform [12]. Subjects were mainly undergraduate and graduate students from a variety of majors at the University of Adelaide recruited using the online recruitment system ORSEE [13]. Approximately thirty percent of participants were majoring in commerce, economics or finance.

Of the remaining seventy percent, most participants were from an engineering, law or science background. Fifty percent of participants were female. No subject participated in more than one session of the experiment. On average, sessions lasted 90 minutes including initial instruction and payment of subjects. Subjects earned an average of $AUD 24.[9]

4. Results

4.1. Average Contributions

Figure 2 illustrates the time path of individual contributions averaged across subjects in each of the six treatments. The top panels show average contributions by period. Average contributions per sequence of four periods are shown in the lower panels. Table 6 in the Appendix shows the corresponding average contributions by sequence for each treatment along with the overall average. Added to this table are the median and the standard deviation of contributions. We observe substantial differences in contributions between treatments. Average contributions are highest in the Shared High Cost and Costless Pledge treatments (77.81 and 76.98, respectively), followed in turn by the No Commitment (72.21), Shared Low Cost (71.68), Asymmetric Cost (70.21) and Random Matching (53.03).

Mann–Whitney pairwise statistical tests comparing the contributions between treatments yield the results shown in Table 2. The unit of observation is the average contribution of individual subjects per sequence.[10] The results confirm what was seen graphically in Figure 2: contributions are highest in the Costless Pledge and Shared High Cost treatments. There is no statistical difference in contributions between these two treatments overall. Contributions in Shared Low Cost are below those in No Commitment over the first four sequences, but similar over the last four sequences. There is no difference in contributions between the No Commitment and Asymmetric Cost treatment. All treatments with endogenous regrouping feature higher contributions than Random Matching.

Observation 1. Overall cooperation is highest in the Costless Pledge and Shared High Cost treatments and lowest under Random Matching. In the Asymmetric Cost and No Commitment treatments, contributions are similar, lying

---

[9] Accumulated earnings were converted to Australian dollars at the predetermined exchange rate of $\$E200 = \$AUD1$ at the end of each session.

[10] If we use individual contributions per period, all differences are significant at $p < 0.01$. In the Mann–Whitney U-tests using sequence averages across all periods, we have 240 independent observations for Random Matching, 448 for No Commitment, 464 for Costless Pledge, 480 for Asymmetric Cost, 400 for Shared High Cost and 480 for Shared High Cost. Dividing the sequences into the first- and second-half of the treatment means the number of independent observations by treatment is divided by a factor of two.

above Random Matching, but below the best performing treatments. In the Shared Low Cost treatment, cooperation gradually increases to a level equal to the No Commitment treatment.

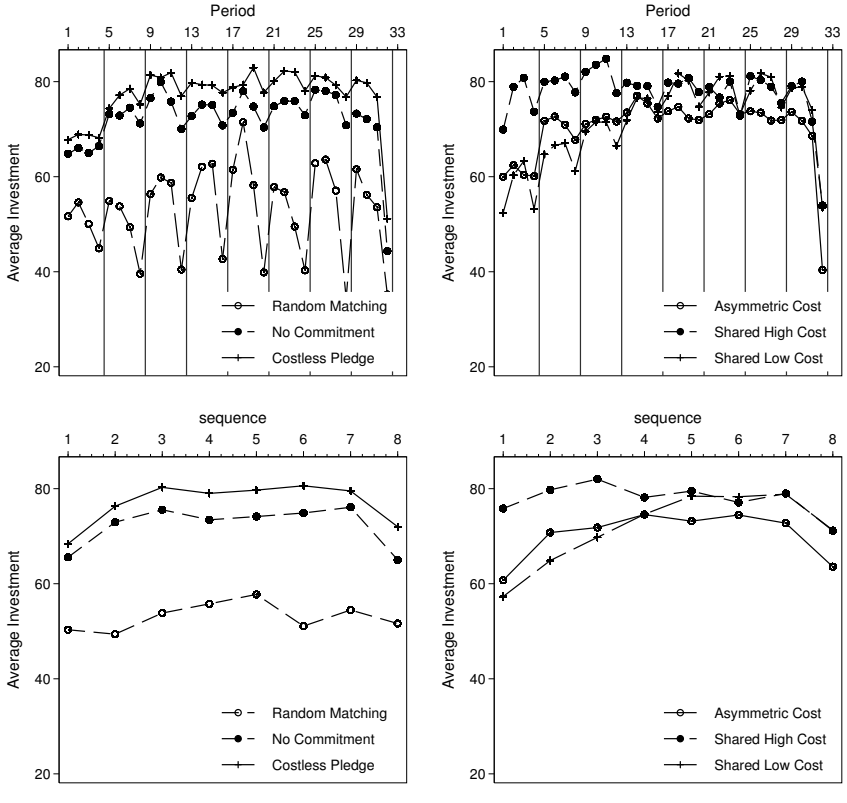


Figure 2. Average individual contributions by treatment.

Higher contributions do not necessarily translate one-to-one with higher welfare because of the payoff reducing costs associated with partnership dissolution in some treatments. To assess whether dissolution costs effect the welfare ranking, we run the same Mann–Whitney U-tests as above on average sequence profits for each treatment. The results of these tests yield the efficiency rankings displayed in Table 3. These rankings highlight that Shared High Cost and Costless Pledge yield equal gains in efficiency and are the only treatments that outperform the No Commitment treatment over the duration of the experiment.

Table 2. Mann–Whitney U-test results: Overall Efficiency.

	No Commitment	Costless Pledge	Asymmetric Cost	Shared Low Cost	Shared High Cost
All Sequences					
Random Matching	8.463 ***	10.587 ***	7.790 ***	8.440 ***	10.921 ***
No Commitment	-	3.055 ***	0.064	0.762	3.709 ***
Costless Pledge	-	-	2.899 ***	4.178 ***	0.585
Asymmetric Cost	-	-	-	1.144	3.547 ***
Shared Low Cost	-	-	-	-	5.096 ***
Sequences 1–4					
Random Matching	6.071 ***	7.418 ***	5.569 ***	4.731 ***	8.276 ***
No Commitment	-	1.539	0.218	2.241 **	2.822 ***
Costless Pledge	-	-	1.735 *	4.196 ***	1.430
Asymmetric Cost	-	-	-	2.197 **	3.048 ***
Shared Low Cost	-	-	-	-	5.666 ***
Sequences 5–8					
Random Matching	5.893 ***	7.630 ***	5.489 ***	7.212 ***	7.184 ***
No Commitment	-	2.755 ***	0.291	1.287	2.411 **
Costless Pledge	-	-	2.444 **	1.898 *	0.450
Asymmetric Cost	-	-	-	0.588	2.090 **
Shared Low Cost	-	-	-	-	1.664 *

Notes: The absolute value of the z-statistic reported is for the Mann–Whitney U-test that average individual contributions per sequence (of four periods) are equal between treatments. Significance levels: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3. Treatment efficiency ranking.

All Sequences		Sequences 1–4		Sequences 5–8	
Ranking	**Treatment**	**Ranking**	**Treatment**	**Ranking**	**Treatment**
1	Shared High Cost	1	Shared High Cost	1	Shared High Cost
1	Costless Pledge	2	Costless Pledge	1	Costless Pledge
3	Shared Low Cost	2	Asymmetric Cost	3	Shared Low Cost
3	Asymmetric Cost	2	No Commitment	3	Asymmetric Cost
3	No Commitment	5	Shared Low Cost	3	No Commitment
6	Random Matching	6	Random Matching	6	Random Matching

Notes: Ranking based on Mann–Whitney U-test results that average individual profits per sequence (of four periods) are equal between treatments.

In the remaining subsections, we dig deeper into the data to explain how these efficiency differences arise. Next, we look at cooperation levels for committed and uncommitted partners in each treatment. We then turn to analyse differences in the proportion of committed partnerships by treatment. Finally, we investigate the dissolution decisions of subjects.

4.2. Commitment Status and Contributions

Figure 3 disaggregates average contribution profiles for committed and uncommitted subjects in each experimental treatment by sequence. In each panel, the average contribution of subjects in the No Commitment treatment

is included as a reference. Table 7 in the Appendix presents the average contributions by treatment and sequence corresponding to Figure 3. The figure clearly documents that committed partnerships are associated with higher average contributions than uncommitted partnerships for all treatments, except Asymmetric Cost. Contributions of committed partners also lie above the No Commitment benchmark.

To tease out the relative effectiveness of commitment on increasing contributions of committed subjects between treatments, we run a series of Mann–Whitney U-tests. The unit of observation is the average contribution level of a subject within a sequence. Table 4 presents the results for all sequences and those disaggregated into the first and last four sequences. The results show that contributions of committed subjects are similar in Costless Pledge, Shared Low Cost and shared hight cost. We cannot reject the null hypothesis of equal average contributions of committed subjects in all pairwise tests between Costless Pledge, Shared Low Cost and Shared High Cost. This confirms what we see graphically in Figure 3. We can also see that the Asymmetric Cost treatment is the least successful of all treatments in getting committed subjects to overcome the social dilemma.

Observation 2. Contributions of committed partners are higher than a No Commitment benchmark when there are no costs to dissolve committed partnerships, or these costs fall equally. There is no treatment difference in the contributions of committed subjects between Costless Pledge, Shared Low Cost and Shared High Cost. Committed subjects in Asymmetric Cost perform the worst, with contributions lower than subjects in the No Commitment reference.

This result provides a first glance at the mechanisms driving the treatment differences in overall cooperation and efficiency. The gain in efficiency documented in Observation 1 in Shared High Cost, and Costless Pledge is driven by the higher contributions of committed subjects. However, this increase in cooperation alone cannot fully explain the efficiency result, because it does give rise to the rankings in Table 3. In particular, it cannot explain why Shared Low Cost does not perform better than No Commitment in terms of treatment level welfare. We now look to commitment rates and dissolution decisions to provide further evidence of the mechanisms at work.

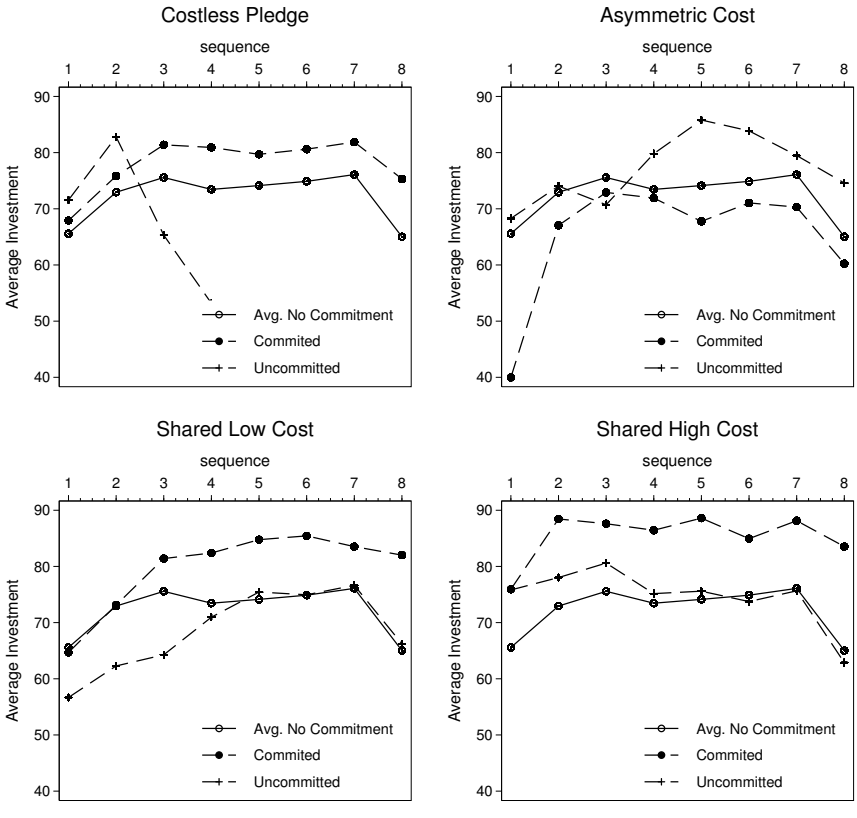


Figure 3. Average contributions by commitment status and sequence.

Table 4. Differences in contributions of committed subjects.

	Asymmetric Cost	**Shared Low Cost**	**Shared High Cost**
All Sequences			
Costless Pledge	4.243 ***	0.505	2.436 ***
Asymmetric Cost	-	3.544 ***	5.152 ***
Shared Low Cost	-	-	1.871 *
Sequences 1–4			
Costless Pledge	2.381 ***	0.049	1.712
Asymmetric Cost	-	1.679 *	3.018 ***
Shared Low Cost	-	-	1.593
Sequences 5–8			
Costless Pledge	3.745 ***	0.539	1.572
Asymmetric Cost	-	3.261 ***	4.232 ***
Shared Low Cost	-	-	1.001

Notes: The absolute value of the z-statistic reported is for the Mann–Whitney U-test that average individual contributions per sequence (of four periods) are equal between treatments. Significance levels: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

4.3. Proportion of Committed Partnerships

The proportion of committed partnerships across treatments and sequences is shown in Figure 4. The proportion is highest in the Costless Pledge treatment, staying above eighty percent in all sequences. A test of proportions confirms that the commitment rate is higher in Costless Pledge than any other treatment ($p < 0.01$ for all pairwise comparisons). In Sequences 5 and 6, every partnership in Costless Pledge is committed.

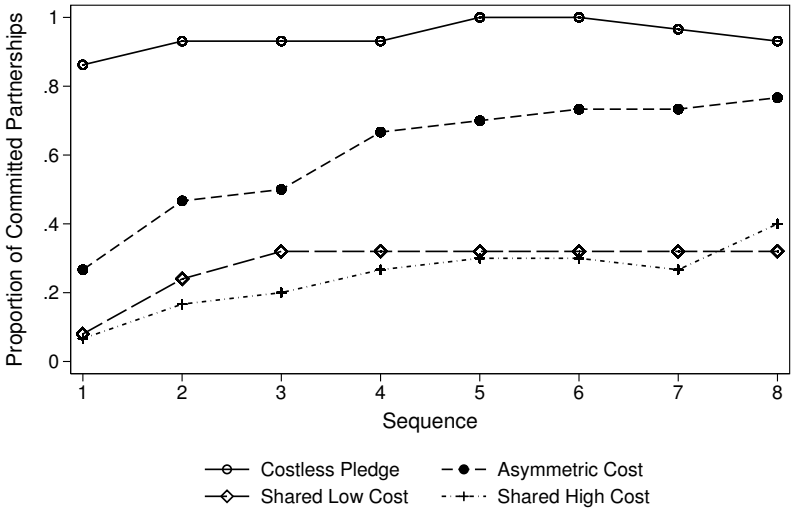


Figure 4. Proportion of committed partnerships by treatment.

Variations in the distribution of dissolution costs between partners leads to significantly different proportions of committed partnerships. The proportion is higher in Asymmetric Cost, where only the partner choosing to dissolve a committed partnership pays a cost, compared to treatments where the dissolution costs are equally shared between partners ($p < 0.01$ for pairwise tests of proportions compared to Shared Low Cost and Shared High Cost respectively). There is no significant difference between Shared Low Cost and Shared High Cost, where the distribution of costs remains constant.

Observation 3. Partnerships are most likely to be committed when there are no costs to dissolving unsuccessful partnerships. The distribution of dissolution costs matter; there are more committed partnerships when costs are borne only by partners who instigate dissolution compared to when costs are equally shared by both partners.

Observations 2 and 3 suggest that treatment differences in overall efficiency documented in Table 3 are driven in part by composition effects. Whilst average contributions were similar in Costless Pledge and Shared Low Cost, there are less committed partnerships in Shared Low Cost. The low proportion of high contributing committed partnerships means that these cooperation gains did not translate into treatment level welfare gains in Shared Low Cost. At the same time, the high proportion of committed partnerships in Costless Pledge mean the cooperation increase did translate into treatment level welfare gains. This gain was similar in magnitude to that in Shared High Cost.

This high proportion of committed partnerships in Costless Pledge combined with the gains in cooperation in committed partnerships provide evidence that costless pledges serve to signal or reassure a partner of cooperative intentions. Furthermore, these pledges are effective in leading to higher cooperation levels in committed partnerships because partners act on these intentions when partnership formation is endogenous. This stands in contrast to [14], who show that structured pledges of commitment are ineffective at increasing cooperation under exogenous regrouping protocol.

4.4. Partnership Dissolution

We now turn to analyse partnership dissolution decisions.[11] Dissolution rates by treatment and sequence are plotted in Figure 5. The left panel documents the fraction of partnerships that are dissolved in the No Commitment and Costless Pledge treatments. Dissolution rates for treatments where dissolving a committed partnership is costly are shown in the right panel. We can see that the introduction of commitment opportunities decreases the aggregate dissolution rate.[12] These treatment differences in dissolution rates documented in Figure 5 are a combination of two effects: the direct effect of commitment and an indirect effect through changes in contributions. To disentangle the two effects, we estimate a random effects linear

---

[11] In the experiment, we asked subjects to type free-form messages after dissolving their partnership to explain the reason for their decision. Nearly all subjects respond that they choose to dissolve the partnership because investments were not high enough. This response is the most common across treatments. Some subjects also respond that because they proposed to commit and the other partner did not, they chose to dissolve

[12] Pairwise comparisons between No Commitment and both Costless Pledge and Shared Low Cost reveal treatment differences at the 10% significance level (test of proportions, $p < 0.09$ for shared cost and $p < 0.08$ for Costless Pledge). Treatment differences are significant at the 1% level between No Commitment and both Asymmetric Cost and Shared High Cost, respectively. Across treatments where subject can choose to commit, the partnership dissolution rate in Asymmetric Cost is lower than in all other treatments ($p < 0.01$ for all pairwise comparisons). There is no significant difference in overall dissolution rates between the Costless Pledge, Shared Low Cost and Shared High Cost treatments ($p > 0.1$ for all pairwise comparisons).

probability model. The dependent variable in the regression is the dissolution decision of each subject in a regrouping phase.[13] The regression coefficients are reported in Table 5.

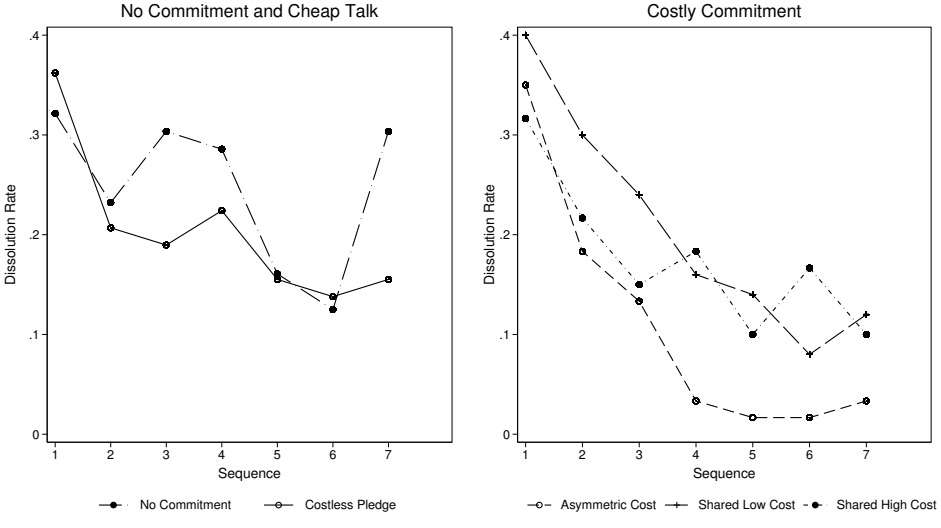


Figure 5. Partnership dissolution rates by treatment and sequence.

Looking first at the coefficients on treatment indicators, only the Costless Pledge coefficient is significantly different from zero. In all other treatments, the introduction of commitment opportunities has no effect on the dissolution decisions of uncommitted subjects. Uncommitted subjects in Costless Pledge are more likely to dissolve because these (few) uncommitted partnerships were unsuccessful at achieving cooperation, so subjects were dissolving their partnerships, frequently anticipating re-matching to a more cooperative partner. The negative coefficients on the commitment-treatment interactions show subjects in committed partnerships are less likely to dissolve their partnership than their uncommitted counterparts in the same treatment. Compared to the reference treatment, No Commitment, subjects in committed partnerships that are costly to dissolve are less likely to break up their partnership. This is not true for Costless Pledge, where committed

[13] Treatment indicators are included as independent variables, as are interactions between commitment status and the treatment indicator. The reference treatment is No Commitment, where pledges of commitment were not available to subjects. We add the minimum and maximum contribution of a subject and their partner over a sequence as controls for contributions to isolate the direct effect of commitment. Sequence indicators are added to control for time effects and subject demographics, such as age, gender and study major, are added as additional controls.

subjects are equally to dissolve as subjects in No Commitment. The distribution of dissolution costs plays an important role in the dissolution decision. For a given contribution profile, committed subjects are 10.7% less likely to dissolve their partnership in Asymmetric Cost compared to Shared Low Cost. This difference is statistically significant at the 5% level. There is no difference in dissolution rates between Shared Low Cost and Shared High Cost.

Table 5. Random effects linear probability model for partnership dissolution.

Pr(Dissolve)	**(1) RELPM**	**(2) RE LPM**	**(3) RE LPM**
Treatment Indicators			
Costless Pledge	0.198 ***	0.165 **	0.169 **
	(0.0740)	(0.0738)	(0.0745)
Asymmetric Cost	−0.000536	−0.0216	−0.0248
	(0.0372)	(0.0374)	(0.0392)
Shared High Cost	−0.00159	−0.00807	−0.00617
	(0.0324)	(0.0324)	(0.0334)
Shared Low Cost	−0.0292	−0.0320	−0.0347
	(0.0345)	(0.0345)	(0.0354)
No Commitment		reference treatment	
Commitment Effects			
Commit × Costless Pledge	−0.222 ***	−0.188 ***	−0.192 ***
	(0.0710)	(0.0708)	(0.0710)
Commit × Asymmetric Cost	−0.252 ***	−0.216 ***	−0.205 ***
	(0.0359)	(0.0365)	(0.0372)
Commit × Shared High Cost	−0.131 ***	−0.107 **	−0.106 **
	(0.0450)	(0.0451)	(0.0457)
Commit × Shared Low Cost	−0.116 ***	−0.101 **	−0.0953 **
	(0.0447)	(0.0446)	(0.0451)
Contributions controls			
Own Max/10	0.0192 ***	0.0183 ***	0.0176 ***
	(0.00526)	(0.00524)	(0.00530)
Own Min/10	−0.0229 ***	−0.0212***	−0.0215 ***
	(0.00354)	(0.00354)	(0.00356)
Partner Max/10	−0.0323 ***	−0.0334 ***	−0.0334 ***
	(0.00510)	(0.00507)	(0.00510)
Partner Min/10	−0.0286 ***	−0.0270 ***	−0.0264 ***
	(0.00347)	(0.00347)	(0.00349)
Constant	0.676 ***	0.742 ***	0.716 ***
	(0.0396)	(0.0426)	(0.0550)
Sequence Dummies		YES	YES
Demographics			YES
% Pred $\in [0,1]$	0.88	0.78	0.78
σ_u	0.121	0.121	0.124
σ_e	0.298	0.296	0.296
R^2	0.323	0.332	0.336
Observations	1988	1988	1988
Number of Subjects	284	284	284

Standard errors in parentheses, clustered at the session level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Observation 4. The introduction of costly pledges of commitment decreases the likelihood that committed partners choose to dissolve their partnership, but have no effect on uncommitted partnerships. Dissolution rates are lowest when

the cost burden falls solely on the partner choosing to dissolve, but are no different when we scale up the cost holding the distribution constant. Costless pledges have no effect on dissolution decisions of committed subjects compared to a setting where commitment opportunities are not available.

The decline in the likelihood of dissolution in Asymmetric Cost illustrates the potential mechanism driving the ineffectiveness of commitment to increase cooperation in the treatment. Upon committing to a partnership, a subject (correctly) anticipates that his or her partner is unwilling to choose to dissolve the partnership and bear the high costs alone, perceiving the cost too high or the distribution unfair. In response, the subject begins to free-ride on his or her partner, decreasing contributions. If his or her partner is a conditional cooperator, he or she in turn responds by decreasing cooperation [15,16]. As a result, the partnership becomes trapped in a low cooperation state with both partners reluctant to dissolve. This suggests that the relatively large presence of committed partnerships in the treatment explains the lower treatment efficiency. Despite the low contributing committed partnerships, treatment efficiency in Asymmetric Cost does not fall below the No Commitment benchmark. In this treatment, uncommitted partnerships use the threat of costless dissolution to maintain high levels of cooperation, mitigating the decline.

The regression results also highlight a difference in the mechanisms driving the cooperation gains in Costless Pledge and Shared High Cost. In the Costless Pledge treatment, we have documented evidence that pledges can be used to signalling cooperative intentions. Despite this signal and the rise in contributions, we do not see a decrease in partnership dissolution in this treatment relative to No Commitment. This contrasts with Shared High Cost, where we see a substantial decline in breakups of committed partnerships. We interpret this result as an alternative mechanism driving welfare gains in this treatment. Although fewer subjects choose to commit in Shared High Cost, those that enter committed partnerships remain with their partner for longer. Cooperation is sustained in these partnerships, unlike Asymmetric Cost, because partners respond to the threat of credible dissolution by keeping contributions high.

5. Conclusions

We report on a set of experiments that were designed to investigate if pledges of commitment enhance cooperation in endogenously-formed partnerships. The design isolates the role of the pledge of commitment separately from the effect of costs incurred by partners when their partnership is dissolved. Our results show that pledges by themselves can yield increases in cooperation within committed partnerships and are welfare improving. In the Costless Pledge treatment, more than 80 percent of all partnerships feature mutual pledges of commitment. These

partnerships have higher rates of cooperation than a benchmark treatment where commitment opportunities are removed. The results suggest that pledges of commitment serve to signal cooperative intentions. Furthermore, once subjects are committed, they coordinate on higher contributions.

A specific focus of the experiment was to analyse how the size and distribution of costs attached to dissolving committed partnerships influence cooperation. In this dimension, our results emphasize the importance of costs being split between subjects when partnerships are dissolved. In two treatments where the costs were equally split between partners, we documented higher cooperation levels in committed partnerships. In these cases, we found a similar increase in cooperation among committed partnerships compared to the Costless Pledge regime. For the Shared High Cost treatment, this translated into welfare gains of a similar magnitude to Costless Pledge. However, the adverse effect the introduction of commitment had on uncommitted subjects in early sequences plus the small number of committed partnerships meant no welfare gains were present in Shared Low Cost. In Asymmetric Cost, where costs fell solely on the instigator of the breakup, cooperation levels fell. In this treatment, subjects anticipated the reluctance of their partner to dissolve the relationship and to bear all of the financial cost by free-riding.

Our results show that pledges of commitment by themselves, and with high and equally-shared dissolution costs, increase welfare in endogenously-formed partnerships. When the total cost burden of equally-shared costs is too low or costs fall only on instigators of breakups, we find no effect on treatment level welfare. This stands in contrast to the previously-studied notions of costly sanctions [17] and monetary punishment [18–20] that have mixed effects on welfare [21] and can be welfare decreasing. Furthermore, there is limited evidence that costly sanctioning and punishment schemes are widely used outside laboratory environments [22], unlike voluntary association and notions of commitment to a partnership. We document that group formation when combined with pledges of commitment can increase cooperation and welfare to a level similar to the best performing punishment treatments [23]. This increase in cooperation comes at both zero cost with pure pledges and at a one-to-one ratio when dissolution costs are equally shared, in contrast to the most effective punishment schemes that rely on ratios above one-to-three.

Acknowledgments: We thank the Editor, Ananish Chaudhuri, and two anonymous referees for their comments that improved the manuscript. We also thank attendees at the 2011 Australasian Meeting of the Econometric Society, 2011 Australian Conference of Economists, 2011 Australia and New Zealand Workshop for Experimental Economics, Ben Greiner, Changxia Ke and Virginie Masson, for useful comments on earlier versions of this paper. We gratefully acknowledge Mickey Chan and Nok Shatragram, who assisted in programming and running sessions of the experiment.

Author Contributions: Lachlan Deer contributed to the design of the experiment, ran the experimental sessions, performed the statistical analysis, interpreted results and wrote the manuscript. Ralph-C. Bayer contributed to the design of the experiment, interpreted results and wrote the manuscript.

Conflicts of Interest: Conflicts of Interest: The authors declare no conflict of interest.

Appendix

Appendix Tables: Contribution Behaviour

Table 6. Summary of contribution behaviour by sequence and treatment.

		Sequence								
Treatment		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
Random Matching	Mean contr.	50.32	49.39	53.83	55.77	57.76	51.09	54.47	51.63	53.03
	Median contr.	50.00	50.00	50.00	59.50	60.00	50.00	50.00	50.00	50.00
	SD	33.00	37.96	36.39	34.47	36.69	36.37	36.66	38.86	36.31
No Commitment	Mean contr.	65.58	72.95	75.57	73.46	74.14	74.88	76.09	65.03	72.21
	Median contr.	70.00	85.00	82.00	80.00	80.00	80.00	90.50	72.50	80.00
	SD	34.69	30.84	28.35	27.56	28.78	28.28	28.87	36.48	30.86
Costless Pledge	Mean contr.	68.41	76.31	80.28	79.01	79.69	80.61	79.53	71.99	76.98
	Median contr.	80.00	90.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	SD	33.73	30.57	26.22	29.51	29.94	29.87	31.66	38.40	31.64
Asymmetric Cost	Mean contr.	60.74	70.77	71.82	74.54	73.18	74.47	72.75	63.59	70.21
	Median contr.	60.00	80.00	87.50	95.00	100.00	100.00	100.00	82.50	90
	SD	35.53	34.39	34.99	32.87	34.07	34.08	37.54	41.65	35.99
Shared Low Cost	Mean contr.	57.31	64.89	69.76	74.62	78.43	78.31	78.85	71.28	71.68
	Median contr.	55.00	70.00	80.00	80.00	90.00	90.0	90.00	90.00	80
	SD	30.10	30.68	30.87	28.35	26.43	26.87	25.62	34.06	30.04
Shared High Cost	Mean contr.	75.83	79.76	82.00	78.17	79.50	77.10	78.99	71.15	77.81
	Median contr.	90.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	SD	28.73	30.36	29.58	33.09	30.58	32.84	30.55	38.39	31.98

Table 7. Summary of committed subject's contribution behaviour by sequence and treatment.

			Sequence								
Treatment			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
No Commitment	All Subjects	Mean Contr.	65.58	72.95	75.57	73.46	74.14	74.88	76.09	65.03	72.21
		SD	34.69	30.85	28.35	27.56	28.76	28.28	28.87	36.48	30.86
Costless Pledge	Committed	Mean Contr.	67.91	75.83	81.39	80.93	79.70	80.61	81.88	75.33	78.10
		SD	33.80	30.87	25.24	28.43	29.94	29.87	29.54	36.52	30.89
	Uncommitted	Mean Contr.	71.56	82.81	65.31	53.19	-	-	13.75	26.88	58.18
		SD	33.64	26.01	34.57	32.48	-	-	13.02	35.58	37.81
Asymmetric Cost	Committed	Mean Contr.	39.98	67.04	72.92	71.91	67.76	71.05	70.30	60.24	66.96
		SD	34.68	35.52	32.25	34.87	35.81	35.72	38.37	42.16	37.40
	Uncommitted	Mean Contr.	68.28	74.04	70.73	79.81	85.81	83.88	79.48	74.59	75.23
		SD	32.80	33.16	37.63	27.91	25.71	27.21	34.55	38.24	33.15
Shared Low Cost	Committed	Mean Contr.	64.69	73.13	81.41	82.39	84.77	85.44	83.53	82.02	81.51
		SD	23.49	23.56	21.87	22.63	22.01	19.62	20.19	25.25	22.54
	Uncommitted	Mean Contr.	56.67	62.28	64.27	70.96	75.44	74.96	76.65	66.23	67.86
		SD	30.57	32.25	32.98	30.06	27.85	29.14	27.61	36.50	31.68
Shared High Cost	Committed	Mean Contr.	75.94	88.45	87.60	86.44	88.60	84.94	88.14	83.54	86.11
		SD	21.23	18.04	17.83	23.03	18.98	24.39	21.50	33.59	24.18
	Uncommitted	Mean Contr.	75.82	78.03	80.60	75.16	75.60	73.75	75.66	62.88	75.10
		SD	29.23	32.01	31.73	35.64	33.68	35.39	32.66	39.26	33.71

References

1. Isaac, R.; McCue, K.F.; Plott, C.R. Public goods provision in an experimental environment. *J. Public Econ.* **1985**, *26*, 51–74.
2. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83.
3. Ledyard, J.O. Public goods: A survey of experimental research. In *The Handbook of Experimental Economics*; Kagel, J.H., Roth, A.E., Eds.; Princeton University Press: Princeton, NJ, USA, 1995.
4. Ahn, T.; Isaac, R.M.; Salmon, T.C. Coming and going: Experiments on endogenous group sizes for excludable public goods. *J. Public Econ.* **2009**, *93*, 336–351.
5. Ahn, T.K.; Isaac, R.M.; Salmon, T.C. Endogenous group formation. *J. Public Econ. Theory* **2008**, *10*, 171–194.
6. Coricelli, G.; Fehr, D.; Fellner, G. Partner selection in public goods experiments. *J. Confl. Resolut.* **2004**, *48*, 356–378.
7. Page, T.; Putterman, L.; Unel, B. Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *Econ. J.* **2005**, *115*, 1032–1053.
8. Bayer, R.C. Cooperation in partnerships: The role of breakups and reputation. *J. Inst. Theor. Econ.* **2005**, in press.
9. Hauk, E.; Nagel, R. Choice of partners in multiple two-person prisoners' dilemma games: An experimental study. *J. Confl. Resolut.* **2001**, *45*, 770–793.
10. Gale, D.; Shapley, L.S. College admissions and the stability of marriage. *Am. Math. Mon.* **1962**, *69*, 9–15.

11. Bayer, R.C. *Cooperation in Partnerships: The Role of Breakups and Reputation*; School of Economics Working Papers 2011–22; University of Adelaide, School of Economics: Adelaide, Australia, 2011.
12. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
13. Greiner, B. Subject pool recruitment procedures: Organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* **2015**, *1*, 114–125.
14. Camera, G.; Casari, M.; Bigoni, M. *Communication, Commitment, and Deception in Social Dilemmas: Experimental Evidence*; Purdue University Economics Working Papers 1236; Purdue University, Department of Economics: West Lafayette, IN, USA, 2010.
15. Fischbacher, U.; Gaechter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **2010**, *100*, 541–556.
16. Fischbacher, U.; Gaechter, S.; Fehr, E. Are people conditionally cooperative? evidence from a public goods experiment. *Econ. Lett.* **2001**, *71*, 397–404.
17. Masclet, D.; Noussair, C.; Tucker, S.; Villeval, M.C. Monetary and nonmonetary punishment in the voluntary contributions mechanism. *Am. Econ. Rev.* **2003**, *93*, 366–380.
18. Fehr, E.; Gaechter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **2000**, *90*, 980–994.
19. Fehr, E.; Gaechter, S. Altruistic punishment in humans. *Nature* **2002**, *415*, 137–140.
20. Ostrom, E.; Walker, J.; Gardner, R. Covenants with and without a sword: Self-governance is possible. *Am. Political Sci. Rev.* **1992**, *86*, 404–417.
21. Tan, F. Punishment in a linear public good game with productivity heterogeneity. *De Economist* **2008**, *156*, 269–293.
22. Guala, F.; Mittone, L. Experiments in economics: External validity and the robustness of phenomena. *J. Econ. Methodol.* **2005**, *12*, 495–515.
23. Nikiforakis, N.; Normann, H.T. A comparative statics analysis of punishment in public-good experiments. *Exp. Econ.* **2008**, *11*, 358–369.

Alleviation and Sanctions in Social Dilemma Games

Michalis Drouvelis

Abstract: This paper reports an experiment which compares behaviour in two punishment regimes: (i) a standard public goods game with punishment in which subjects are given the opportunity to punish other group members (democratic punishment regime) and (ii) a public goods game environment where all group members exogenously experience an automatic reduction of their income (irrespective of their behaviour) and are given the opportunity to alleviate the automatic penalty (undemocratic punishment regime). We employ a within-subjects design where subjects experience both environments and control for order effects by alternating their sequence. Our findings indicate that average contributions and earnings in the undemocratic punishment environment are significantly lower relative to the standard public goods game with punishment. We also observe that in the undemocratic environment average contributions decay over time only when subjects have experienced the standard public goods game with punishment. As a result, alleviation is significantly less when subjects have experienced the standard public goods game with punishment compared to when they do not have such experience. However, the assignment of punishment is robust irrespective of the order in which the games are played.

Reprinted from *Games*. Cite as: Drouvelis, M. Alleviation and Sanctions in Social Dilemma Games. *Games* **2015**, *6*, 368–380.

1. Introduction

A central theme in the behavioural sciences is the examination of the ability of punishment regimes, where sanctioning is assigned endogenously by group members to their peers, to regulate self-interested behaviour in social dilemma games (as in [1,2]). We refer to these environments as "democratic punishment regimes". Numerous laboratory experiments on public goods have shown that such democratic punishment regimes foster high cooperation rates (for recent overviews, see [3,4]). However, this cooperation enhancing effect has been found to be sensitive to a number of factors, such as low effectiveness of punishment (see [5–8]), second-round punishment opportunities (see [9–11]), and antisocial punishment (see [4,12,13]). The evidence from these experiments suggests that, under certain conditions, peer punishment can have little or negative effect on cooperation. The aim of this paper is to explore experimentally how subjects respond in a democratic punishment regime when they have experience with an undemocratic punishment regime and *vice versa*.

To generate an undemocratic punishment regime, we propose a variant of the standard punishment game, which we refer to as the "default punishment game". In this game, group members participate in a two-stage game. In the first stage, they are engaged in a standard linear public goods game, in which they have to decide how much of their initial endowment they are willing to contribute to the public good (see [14]). The game is parameterized so that it captures the tension between personal interests and social benefits. At the beginning of the second stage, we introduce a new element: each group member's income is *exogenously* reduced by a fixed automatic penalty imposed on them—the default punishment. This implies that all group members unconditionally receive a decrease in their monetary income, irrespective of their first stage behaviour. During the second stage, after contribution decisions have been anonymously revealed, subjects are given the opportunity to alleviate the exogenous default punishment of others at some cost to themselves. Our environment has the interesting feature that the fixed income reduction does not depend on the individual behaviour. The automatic penalty is undemocratic in the sense that it is exogenously implemented and the sanctioning mechanism is not being any more social as it cannot identify individual defectors or cooperators.[1] We thus assume that our punishment regime generates a less democratic environment in relation to the standard punishment game where punishment is meted out endogenously by group members.

Our paper is also related to the literature examining behaviour in situations where subjects are given the possibility (typically using voting procedures) to choose the public goods game environment they would like to interact in, as well as the punishment rules governing these environments (e.g., [16–19]).[2] In sum, this literature suggests that letting subjects democratically choose which regime they prefer positively affects cooperative behaviour.

In our paper, the automatic penalty is exogenously assigned to all group members (including those with cooperative attitudes) and it is in this respect that our punishment regime is undemocratic. By having subjects experience an undemocratic punishment environment before they play the standard punishment game, we can assess whether our undemocratic environment impacts on individuals' expectations

---

[1] The effects of blind punishment on public good provision have been recently studied by [15]. In particular, they use a punishment mechanism based on random exclusions. Yet, their design is distinctive to ours in two main respects concerning the rules governing the implementation of punishment. First, their punishment scheme is still social and pursues a collective goal. Good teams were never punished under their sanctioning system. Second, unlike our experiment where subjects are given the opportunity to alleviate the automatic penalty, in their design there is no second stage in which individuals can correct the unfairness of the blind punishment. Their findings suggest that random exclusions generate more public good provision (compared to a standard public good game without punishment) and promote efficiency in a significant way.

[2] For a review of the experimental literature on the workings of democratic institutions, see [20].

of how punishment works. We also reverse the order of the two games. This allows us to test whether, and if so how, experience of a democratic punishment regime affects behaviour in the undemocratic punishment regime.

Our findings indicate that more pronounced detrimental effects on behaviour are present in the undemocratic punishment regime when subjects have experienced the democratic punishment regime. However, this is not the case for the democratic punishment regime where behaviour is not affected by the experience (or not) of the undemocratic default punishment game. Overall, we find that performance in the undemocratic punishment regime is worse in relation to the democratic one.

The remainder of the paper is organised as follows. Section 2 presents the design and the procedures of the experiment. Section 3 reports the results and Section 4 concludes.

2. Experimental Design and Procedures

2.1. Experimental Design

To address our research questions, we implemented a within-subjects design which consists of two conditions: the "default punishment condition" (D-condition) and the "standard punishment condition" (S-condition). We alternated the order of the two conditions, enabling us to control for potential order effects. This resulted in including two sequences in our experiment: the DS sequence, in which the D-condition is followed by the S-condition and the SD sequence, in which the S-condition is followed by the D-condition.

In each condition, subjects interacted in a two-stage game for 10 periods. The group composition remained fixed across all 20 periods within a sequence (Partners' matching protocol). The first stage of the game was common to both conditions involving a voluntary contributions mechanism game with linear payoffs. During the first stage, subjects, being randomly assigned to a four-person group, are privately endowed with 20 tokens each and have to decide how many of these to keep for themselves and how many to contribute to a public good (described to subjects as "project"). For each token kept, each subject earns 1 Money Unit (MU); whereas, for each token contributed the return is equal to 0.5 MUs, resulting in a total of 2 MUs for the whole group. Subjects make their decisions simultaneously and in private. At the end of the first stage they are informed about the sum of the contributions to the public good made by the whole group and about their own first stage income. The payoff function for a given subject in the first stage is given by Equation (1):

$$\pi_i^1 = 20 - g_i + 0.5 \cdot \left(g_i + \sum_{j=1}^{3} g_{j \neq i} \right) \tag{1}$$

where g_i ($0 \leqslant g_i \leqslant 20$) denotes the number of tokens contributed to the public good by group member i.

After the first stage has finished, a second stage follows. The structure of the second stage differs between the D-condition and the S-condition. We thus discuss the second stage of each condition in turn.

2.1.1. The D-Condition

At the beginning of the second stage of the D-condition, all group members incur an automatic penalty irrespective of their first stage contributions, which we refer to "default punishment".[3] In this stage, a group member can see the profile of contributions of the other three group members and is given the opportunity to alleviate the automatic penalty of other group members. Alleviating the automatic penalty is costly for the person who alleviates, but beneficial for the person receiving the alleviation. Subjects could assign between 0 and 2 adjustment points to each other group member. Assignment of adjustment points costs the person who alleviates 1 MU; whereas the impact of assigning one adjustment point is equal to 3 MUs.

In our experiment, the automatic penalty was set equal to 10 MUs. We did so for two reasons. First, complete alleviation of the automatic penalty was possible only if the majority of the group members decided to assign adjustment points. Since each group member can assign up to 2 adjustment points, with each point decreasing the automatic penalty by 3 MUs, the automatic penalty is fully alleviated only if two or more group members assign the total amount of points they control. Second, we did not want to create a situation where subjects would be very likely to end up with substantial losses due to the automatic penalty at the end of the experiment. In this case, subjects would receive a large lump sum payment to cover possible losses in the D-condition, which could affect their behaviour.

The payoff function for a given subject in the D-condition is given by Equation (2):

$$\pi_i = \pi_i^1 - 10 - \sum_{j \neq i} p_{ij} + 3 \cdot \sum_{j \neq i} p_{ji} \tag{2}$$

where π_i^1 denotes the first stage income (as indicated in Equation (1)), p_{ij} denotes the number of adjustment points that group member i assigns to group member j and p_{ji} denotes the number of points assigned by j to i. Note that if a subject received

[3] In the instructions, at the beginning of the second stage, we specify that the default punishment applies to each group member. In particular, we mention "Regardless of contributions, you will also receive an automatic penalty of 10 Money Units." A copy of the instructions used in the experiment is provided in the Supplementary Material.

more adjustment points than the automatic penalty, their income could not increase by this extra amount.

At the end of the second stage, subjects were informed about their own cost of assigning adjustment points, the total number of adjustment points assigned to them, and their earnings. No information about the number of adjustment points received by each group member was available.

2.1.2. The S-Condition

Contrary to the D-condition, the second stage of the S-condition does not include any automatic penalty. After a group member observes the profile of contributions of the other three group members from the first stage, s/he is given the opportunity to decrease each other group members' income by assigning adjustment points. The assignment of adjustment points is costly both for the punisher and the recipient of the punishment. Each adjustment point assigned costs the punisher 1 MU. For each adjustment point, the recipient's income decreases by 3 MUs. As in the D-condition, subjects can assign up to 2 adjustment points. The payoff function for a given subject in the D-condition is given by Equation (3):

$$\pi_i = \pi_i^1 - \sum_{j \neq i} p_{ij} - 3 \cdot \sum_{j \neq i} p_{ji} \tag{3}$$

where π_i^1 denotes the first stage income (as indicated in equation (1)), p_{ij} denotes the number of adjustment points that group member i assigns to group member j and p_{ji} denotes the number of points assigned by j to i.

At the end of the second stage, subjects received the same feedback information as in the D-condition. Note also that, in both conditions, subject-specific reputations cannot develop across periods, since subject i does not have the information to construct a link between individual contributions of subject j across periods. Since subject-specific reputations cannot build up, the possibility that player i assigns adjustment points to player j in period t for contribution decisions made in a previous period from t is ruled out.

Conditional on each group member i being motivated to maximise Equations (2) and (3), the unique subgame perfect equilibrium requires that, in both conditions, each group member free rides completely in the first stage and refrains completely from assigning adjustment points in the second stage.

2.2. Procedures

All sessions took place in April and May 2008 in the Centre for Decision Research and Experimental Economics (CeDEx) lab. Recruitment was conducted via the software ORSEE ([21]) at the University of Nottingham using subjects from

a university-wide pool of registered students. All conditions were computerized and programmed with the software z-Tree ([22]). At the beginning of each sequence, subjects received instructions for the first condition and at the end of it for the second condition. Subjects were informed that the session consists of two conditions in order to reduce the possibility for having wrong expectations about the nature of the experiment. However, they were not told what will happen in the second condition. All participants answered several test questions, concerning the calculation of payoffs for various hypothetical configurations of behaviour. None of the conditions proceeded until every subject had answered these questions correctly. Each of the two sequences was conducted twice, yielding a total of 4 sessions. In the DS sequence 40 subjects participated, resulting in 10 independent observations; while in the SD sequence 36 subjects participated, resulting in 9 independent observations.

At the end of a sequence, subjects were privately paid according to their accumulated earnings from all 20 periods, using an exchange rate of £0.015 per MU. Average earnings per sequence were as follows: £10.45 for the DS sequence and £9.48 for the SD sequence. Sessions lasted, on average, 75 min.

3. Results

3.1. Contribution Levels

Finding 1. (a) Average contributions are significantly lower in the D-condition compared to the S-condition within a given sequence. (b) In the D-condition, average contributions decline over time only when subjects already have experienced the S-condition.

Support. We begin our data analysis by looking at how contribution levels evolved in each sequence. Data are presented as the amount of tokens contributed to the group account. Starting with the DS sequence, we find that subjects' mean contributions were 13 and 15.09 tokens for the D- and the S-conditions, respectively. Regarding the SD sequence, average contributions across all ten periods were 13.29 tokens for the S-condition and 10.18 tokens for the D-condition. Figure 1 shows the evolution of average contribution levels in each sequence separately. In each panel, we report the mean contributions across periods for each condition, the corresponding standard deviation and the p-values from a Wilcoxon signed-rank test for within sequence comparisons.[4]

---

[4] Note that since subjects' matching was fixed across all periods within a sequence, we treat each matching group as the unit of independent observation.

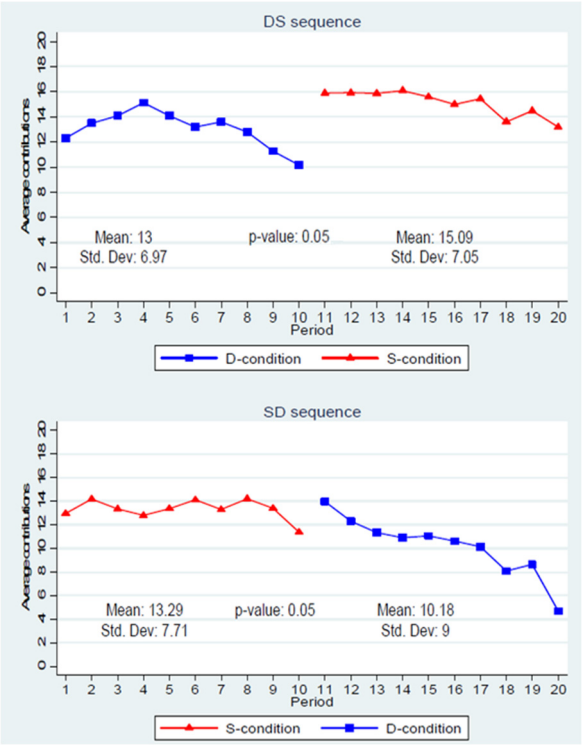


Figure 1. Time series of average contributions for the DS and SD sequences.

Figure 1 shows that contributions over time are stable in both of the S-conditions in either sequences. This is in line with existing findings in the literature of public good games with punishment (e.g., [1]). However, we observe that the contribution patterns of the D-conditions are not stable over time. In the SD sequence, contributions initially increase during the first half of the D-condition and then follow a decaying pattern over the second half. In the DS sequence, we observe that contributions decline over time. By performing a Wilcoxon sign rank test, we find significant within-subject differences across conditions in each sequence. Specifically, in relation to the S-condition, subjects contribute significantly less in the D-condition of the DS sequence ($p = 0.05$) and of the SD sequence ($p = 0.05$).[5]

[5] By contrast, in all but one between-subject comparisons we do not obtain significant differences in contributions ($p > 0.37$). The only exception is the comparison of the S-condition (DS sequence) with the D-condition (SD sequence) where average contributions are significantly higher in the former relative to the latter condition ($p = 0.07$).

Our econometric analysis corroborates these findings when we perform OLS regressions that test for differences across conditions within each sequence, controlling for time trends.[6] We estimate four OLS regressions, in which the dependent variable is the contribution to the group account. In Models (1a) and (1b) we focus on the DS sequence and in Models (2a) and (2b) we focus on the SD sequence. The independent variables for Models (1a) and (2a) consist of a dummy variable for a condition and period, which is equal to the actual number of periods. In Models (1a) and (1b) the dummy variable called "S-condition" is equal to 1 for the S-condition and 0 for the D-condition, whereas in Models (2a) and (2b) the dummy variable called "D-condition" is equal to 1 for the D-condition and 0 for the S-condition. We include the interaction between the treatment dummy and period in Models (1b) and (2b). The regression results are reported in Table 1.

Table 1. Differences in contribution levels within a sequence—Regression results.

	Dependent Variable: Contribution to the Group Account			
	DS Sequence		**SD Sequence**	
	Model 1(a)	Model 1(b)	Model 2(a)	Model 2(b)
S-condition	2.09 ** (0.833)	2.142 ** (0.831)		
D-condition			−3.117 ** (1.063)	0.719 (1.785)
Period	−0.287 (0.190)	−0.282 (0.217)	−0.428 *** (0.088)	−0.080 (0.169)
S-condition × Period		−0.009 (0.130)		
D-condition × Period				−0.697 * (0.318)
Constant	14.579 *** (1.195)	14.553 *** (1.252)	15.647 *** (2.002)	13.730 *** (1.522)
Obs.	800	800	720	720

Note: OLS estimates. Robust standard errors (clustered on independent matching groups) are presented in parentheses. In Models (1a) and (1b) the dummy variable called "S-condition" is equal to 1 for the S-condition and 0 for the D-condition, whereas in Model (2a) and (2b) the dummy variable called "D-condition" is equal to 1 for the D-condition and 0 for the S-condition. The variable "Period" is equal to the actual number of periods. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.

The regression coefficients of Models (1a) and (1b) indicate that subjects contribute significantly more in the S-condition relative to the D-condition of the DS sequence. Additionally, we find no significant time trends for the evolution of contributions in the D-condition as well as in the S-condition (as the variable "Period" and the interaction term between the dummy variable "S-condition" and "Period" are not statistically significant at conventional levels). When we look at Model (2a), we find that contributions are significantly lower in the D-condition compared to the S-condition of the SD sequence. Furthermore, in Model (2b), we find that contributions do not decline over time in the S-condition

[6] We obtain similar results when we estimate linear mixed and Tobit regression models. We report these regressions in Appendix A (see Tables A1 and A2 respectively).

(as the variable "Period" is insignificant). However, the contribution pattern seems to be different in the D-condition (as the interaction term between the treatment dummy variable "D-condition" and "Period" is negative and statistically significant). When we perform a post-estimation test for the variables "Period" and the respective interaction term, we find that contributions are decaying over time in the D-condition ($p = 0.00$).

3.2. Alleviation and Sanctions

Finding 2: (a) Subjects alleviate significantly more in the D-condition of the DS sequence compared to the D-condition of the SD sequence. (b) Subjects do not sanction significantly differently in the S-condition of the DS sequence compared to the S-condition of the SD sequence.

Support. We next turn our attention to how alleviation and sanctions were assigned in the respective conditions. We first examine whether the alleviation functions differ by estimating an ordered probit regression model which controls for factors that are likely to affect alleviation behaviour such as positive and negative deviations. As dependent variable, we include the number of points that player i assigns to player j. The independent variables consist of: "Player j's absolute negative (contribution) deviation", "Player j's positive (contribution) deviation", and the dummy variable called "D-condition (DS sequence)". We also include two interaction terms, which indicate whether the slope of the alleviation function differs with respect to negative and positive deviations across the two D-conditions. Note that all deviations are calculated with respect to the player i's contribution. The variable "Player j's absolute negative (contribution) deviation" is the absolute value of the actual deviation of player j's contribution from player i's contribution, when player j's contribution is below player i's contribution; and zero otherwise. The variable "Player j's positive (contribution) deviation" is constructed in an analogous way. The dummy variable "D-condition (DS sequence)" equals 1 for the D-condition of the DS sequence and 0 for the D-condition of the SD sequence.

Table 2 reveals two important observations. First, the coefficient of the dummy variable "D-condition (DS sequence)" has a positive sign and is statistically significant at the 5% level, implying that the intercept of the alleviation function is different across our two sequences. This implies that subjects in the D-condition of the DS sequence alleviate more points compared to subjects in the D-condition of the SD sequence. Additionally, the coefficients of the interaction term between negative deviations and the dummy variable "D-condition (DS sequence)" is negative and statistically significant at the 1% level, suggesting that the slope of the alleviation function is steeper in the D-condition of the DS sequence compared to the D-condition of the SD sequence. This is not the case for positive deviations where the slope of the alleviation function does not differ across sequences.

We next analyse punishment behaviour by contrasting second stage behaviour in the S-condition of the SD condition versus the S-condition of the DS sequence. To do this, we estimate an ordered probit model which controls for factors that are likely to affect punishment behaviour such as positive and negative deviations. As dependent variable, we include the number of points that player *i* assigned to player *j*. The independent variables consist of: '"Player *j*'s absolute negative (contribution) deviation", "Player *j*'s positive (contribution) deviation", and the dummy variable called "S-condition (SD sequence)". We also include two interaction terms, which indicate whether the slope of the punishment function differs with respect to negative and positive deviations across the two S-conditions. Note that all deviations are calculated with respect to the player *i*'s contribution. The variables "Player *j*'s absolute negative (contribution) deviation" and "Player *j*'s positive (contribution) deviation" are defined as above. The dummy variable "S-condition (SD sequence)" equals 1 for the S-condition of the SD sequence and 0 for the S-condition of the DS sequence.

Table 2. The alleviation function—Regression results.

	Dependent Variable: Alleviation Assigned by Player *i* to player *j*
Player *j*'s absolute negative (contribution) deviation	−0.040 (0.030)
Player *j*'s positive (contribution) deviation	−0.010 (0.021)
D-condition (DS sequence)	0.759 ** (0.334)
D-condition (DS sequence) × Absolute negative deviation	−0.133 *** (0.036)
D-condition (DS sequence) × Positive deviation	−0.039 (0.027)
Obs.	2280

Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on independent matching groups). The variable "Player *j*'s absolute negative (contribution) deviation" is the absolute value of the actual deviation of player *j*'s contribution from player i's contribution, when player *j*'s contribution is below player *i*'s contribution; and zero otherwise. The variable "Player *j*'s positive (contribution) deviation" is constructed in an analogous way. The dummy variable "D-condition (DS sequence)" equals 1 for the D-condition of the DS sequence and 0 for the D-condition of the SD sequence. ** denotes significance at the 5-percent level, and *** at the 1-percent level.

The results from Table 3 indicate that the vertical intercept of the punishment function does not differ across the two sequences, since the coefficient of the dummy variable "S-condition (SD sequence)" is not statistically significant. The coefficients of the interaction terms are also insignificant, implying that the slope of the punishment function with regard both to negative and positive deviations is the same, irrespective of the sequence. The only variable that has a significant effect on the assignment of punishment is the absolute negative deviation from the punisher's contributions: a subject punishes a co-player more, the less the co-player contributes relative to the punisher. This is in line with previous literature from public good games (e.g., [1,23]) indicating that the punishment function is negatively sloped.

Table 3. The punishment function—Regression results.

	Dependent Variable: Punishment Assigned by Player *i* to Player *j*
Player *j*'s absolute negative (contribution) deviation	0.153 *** (0.025)
Player *j*'s positive (contribution) deviation	−0.027 (0.021)
S-condition (SD sequence)	0.181 (0.263)
S-condition (SD sequence) × Absolute negative deviation	0.002 (0.028)
S-condition (SD sequence) × Positive deviation	−0.008 (0.026)
Obs.	2,280

Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on independent matching groups). The variable "Player *j*'s absolute negative (contribution) deviation" is the absolute value of the actual deviation of player *j*'s contribution from player *i*'s contribution, when player *j*'s contribution is below player *i*'s contribution; and zero otherwise. The variable "Player *j*'s positive (contribution) deviation" is constructed in an analogous way. The dummy variable "S-condition (SD sequence)" equals 1 for the S-condition of the SD sequence and 0 for the S-condition of the DS sequence. *** denotes significance at the 1-percent level.

3.3. Average Earnings

Finding 3: Average earnings are significantly lower in the D-condition compared to the S-condition within a given sequence.

Support. In this section we explore how efficiency (as measured by subjects' net average earnings) was determined in each condition. Table 4 presents the average net earnings in each condition and sequence, separately.

Table 4. Average earnings for each condition and sequence.

	D-Condition	**S-Condition**
DS sequence	25.88 (7.80)	31.75 (7.28)
SD sequence	22.16 (10.04)	29.07 (8.83)

Note: Earnings are measured in Money Units (MUs). Standard deviations in parentheses.

In either sequences the S-condition yields higher welfare (31.75 MUs) in the DS sequence and 29.07 MUs in the SD sequence) compared to the D-condition (25.88 MUs in the DS sequence and 22.16 MUs in the SD sequence). By performing a Wilcoxon sign rank test, we test for differences in earnings within a given sequence. We find that the S-condition yields higher welfare than the D-condition both in the DS sequence ($p = 0.01$) and the SD sequence ($p = 0.01$).[7]

[7] We find insignificant differences across conditions in all pairwise between-subject comparisons ($p > 0.22$) except for the comparison of the S-condition (DS sequence) with the D-condition (SD sequence) where average earnings are significantly higher in the former relative to the latter condition ($p = 0.02$).

4. Concluding Remarks

In this paper, we report an experiment examining how an undemocratic public goods game (D-condition) affects behaviour in a standard public goods game with punishment (S-condition) and *vice versa*. Specifically, in our experiment, we introduce an environment where an automatic penalty is meted out exogenously to all group members (unconditionally and by default), irrespective of their prior behaviour. Subjects are then given the opportunity to alleviate the automatic penalty from others by assigning adjustment points to them. We refer to this game as the default punishment game.

Our findings show that average contributions and earnings are significantly lower in the default punishment game compared to the standard punishment game. We also find that contributions in the default punishment game decline over time only after the experience of the standard public goods game with punishment. This decaying pattern of contributions is explained by differences in how alleviation occurs: subjects alleviate other group members significantly less when they first experience the standard public good game with punishment compared to when they have no such experience.

Our results indicate significant order effects in relation to how subjects respond in the less democratic default punishment regime. However, this is not the case for the standard public good game with punishment where the contribution levels remain robust over time and the assignment of punishment is unaffected irrespective of whether subjects have experienced previously the default punishment game or not.

We see at least two avenues for future research. First, investigating the perception of automatic penalty in subject pools with high levels of antisocial punishment may lead to interesting insights on the behavioural consequences of the default punishment game. Second, in the light of previous research (e.g., [5,6]) suggesting that if norm adherence or enforcement becomes more costly, norms are more likely to collapse, it will be of interest to analyse whether and if so, how the size of automatic penalty impacts on the sustainability of social norms.

Acknowledgments: We would like to thank James Cox, Robin Cubitt, Eamonn Ferguson, Simon Gächter, John Hey, the handling editor, Ananish Chaudhuri, and three anonymous reviewers for valuable comments and suggestions. The paper has also benefited from comments by seminar participants at the University of Nottingham and the ESA Meetings in Innsbruck (2009) and in Copenhagen (2010). Financial support from the ESRC (PTA-030-2005-00608) and the University of Nottingham is gratefully acknowledged.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A—Regression Results

Table A1. Differences in contribution levels within a sequence—Regression results.

	Dependent Variable: Contribution to the Group Account			
	DS Sequence		SD Sequence	
	Model 1(a)	Model 1(b)	Model 2(a)	Model 2(b)
S-condition	2.09 ** (0.832)	2.142 ** (0.829)		
D-condition			−3.117 *** (1.062)	0.719 (1.782)
Period	−0.287 (0.190)	−0.282 (0.217)	−0.428 *** (0.088)	−0.080 (0.168)
S-condition × Period		−0.009 (0.130)		
D-condition × Period				−0.697 ** (0.318)
Constant	14.579 *** (1.193)	14.553 *** (1.249)	15.647 *** (2.000)	13.730 *** (1.519)
Obs.	800	800	720	720

Note: Linear mixed model estimates at independent matching group level. Robust standard errors are presented in parentheses. In Models (1a) and (1b) the dummy variable called "S-condition" is equal to 1 for the S-condition and 0 for the D-condition, whereas in Model (2a) and (2b) the dummy variable called "D-condition" is equal to 1 for the D-condition and 0 for the S-condition. The variable "Period" is equal to the actual number of periods. ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.

Table A2. Differences in contribution levels within a sequence—Regression results.

	Dependent Variable: Contribution to the Group Account			
	DS Sequence		SD Sequence	
	Model 1(a)	Model 1(b)	Model 2(a)	Model 2(b)
S-condition	4.676 ** (2.260)	5.395 * (2.864)		
D-condition			−7.313*** (2.192)	4.908 (6.640)
Period	−0.550 (0.436)	−0.488 (0.459)	−1.062*** (0.290)	−0.028 (0.581)
S-condition × Period		−0.131 (0.335)		
D-condition × Period				−2.232 * (1.141)
Constant	18.557 *** (3.216)	18.222 *** (2.864)	24.058 *** (6.023)	18.096 *** (4.100)
Obs.	800	800	720	720

Note: Tobit estimates. Standard errors presented in parentheses (clustered on independent matching groups) are presented in parentheses. In Models (1a) and (1b) the dummy variable called "S-condition" is equal to 1 for the S-condition and 0 for the D-condition, whereas in Model (2a) and (2b) the dummy variable called "D-condition" is equal to 1 for the D-condition and 0 for the S-condition. The variable "Period" is equal to the actual number of periods. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.

References

1. Fehr, E.; Gächter, S. Cooperation and Punishment in Public Goods Experiments. *Am. Econ. Rev.* **2000**, *90*, 980–994.
2. Fehr, E.; Gächter, S. Altruistic Punishment in Humans. *Nature* **2002**, *415*, 137–140.
3. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83.
4. Gächter, S.; Herrmann, B. Reciprocity, culture and human cooperation: Previous insights and a new cultural experiment. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 761–806.

5. Anderson, C.; Putterman, L. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games Econ. Behav.* **2006**, *54*, 1–24.
6. Carpenter, J. The demand for Punishment. *J. Econ. Behav. Organ.* **2007**, *62*, 522–542.
7. Egas, M.; Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B Biol. Sci.* **2008**, *275*, 871–878.
8. Nikiforakis, N.; Normann, H.-T. A comparative static analysis of punishment in public-good experiment. *Exp. Econ.* **2008**, *11*, 358–369.
9. Cinyabuguma, M.; Page, T.; Putterman, L. Can second-order punishment deter perverse punishment? *Exp. Econ.* **2006**, *9*, 265–279.
10. Denant-Boemont, L.; Masclet, D.; Noussair, C. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Econ. Theory* **2007**, *33*, 145–167.
11. Nikiforakis, N. Punishment and Counter-Punishment in Public Goods Game: Can we still govern ourselves? *J. Public Econ.* **2008**, *92*, 91–112.
12. Gächter, S.; Herrmann, B. The Limits of Self-Governance when Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia. *Eur. Econ. Rev.* **2011**, *55*, 193–210.
13. Herrmann, B.; Thöni, C.; Gächter, S. Anticosial punishment across societies. *Science* **2008**, *319*, 1362–1367.
14. Ledyard, J. Public Goods: A Survey of Experimental Research. In *Handbook of Experimental Economics*; Kagel, J.H., Roth, A.E., Eds.; Princeton University Press: Princeton, NJ, USA, 1995; pp. 111–194.
15. Fatas, E.; Morales, A.J.; Ubeda, P. Blind Justice: An experimental analysis of random punishment in team production. *J. Econ. Psychol.* **2010**, *31*, 358–373.
16. Ertan, A.; Page, T.; Putterman, L. Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *Eur. Econ. Rev.* **2009**, *53*, 495–511.
17. Kamei, K.; Putterman, L.; Tyran, J.-R. State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Exp. Econ.* **2015**, *18*, 38–65.
18. Markussen, T.; Putterman, L.; Tyran, J.-R. Self-organization for collective action: an experimental study of voting on sanction regimes. *Rev. Econ. Stud.* **2014**, *81*, 301–324.
19. Putterman, L.; Tyran, J.-R.; Kamei, K. Public Goods and Voting on Formal Sanction Schemes. *J. Public Econ.* **2011**, *95*, 1213–1222.
20. Dal Bó, P. Experimental Evidence on the Workings of Democratic Institutions, Working Paper 2010. Available online: http://www.econ.brown.edu/fac/Pedro_Dal_Bo/institutionschapter.pdf (accessed on 15 September 2015).
21. Greiner, B. Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *J. Econ. Sci. Assoc.* **2015**, *1*, 114–125.
22. Fischbacher, U. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Exp. Econ.* **2007**, *10*, 171–178.
23. Cubitt, R.P.; Drouvelis, M.; Gächter, S. Framing and free riding: Emotional responses and punishment in social dilemma games. *Exp. Econ.* **2011**, *14*, 254–272.

Section 3

Students, Temporary Workers and Co-Op Workers: An Experimental Investigation on Social Preferences

Davide Dragone, Fabio Galeotti and Raimondello Orsini

Abstract: We conduct an artefactual field experiment to compare the individual preferences and propensity to cooperate of three pools of subjects: Undergraduate students, temporary workers and permanent workers. We find that students are more selfish and contribute less than workers. Temporary and permanent contract workers have similar other-regarding preferences and display analogous contribution patterns in an anonymous Public Good Game.

Reprinted from *Games*. Cite as: Dragone, D.; Galeotti, F.; Orsini, R. Students, Temporary Workers and Co-Op Workers: An Experimental Investigation on Social Preferences. *Games* **2015**, *6*, 79–123.

1. Introduction

Undergraduate university students can be easily recruited for experiments and are the typical object of inquiry in the experimental social sciences. One critical issue linked to this admittedly specific population concerns whether the experimental results obtained with students are robust to using other populations, such as workers or older age people [1–4]. To address this issue, the literature has compared the behavior of undergraduate students with different types of workers, such as nurses [5], Chief Executive Officers [6], workers in a publishing distribution warehouse [7], bicycle messengers [8], and clerical workers [9].

Our paper relates to this stream of research and it investigates, using an artefactual field experiment [10] and a sample of subjects living in an Italian region with high social capital, whether and to what extent the other-regarding preferences and cooperative behavior of undergraduate university students differ from those of workers, and how these preferences translate into behavior in a strategic setting. We further study whether there exists a link between the type of worker's contractual arrangement (temporary or permanent contract) and her social preferences and behavior. The answer to the latter question is still an open issue which has not been fully addressed so far. Yet, this line of inquiry deserves investigation, due to the rapid diffusion of temporary employment (*i.e.*, a work situation where an employee is hired for a pre-determined time limit) in many industrialized countries, in particular in countries providing high levels of employment protection [11].

From a firm's perspective, temporary employment reflects the need for a flexible labor demand which can be quickly adjusted to workload fluctuations. It also makes it easier for a firm to replace less productive people with more productive ones, and may as well favor the inflow of innovative ideas. This type of contractual arrangement can however also have negative effects. A high turnover may in fact hinder social cohesion, cooperative spirit, organizational commitment and trust in the workplace, and increase the probability of opportunistic behavior [12]. In addition, workers who compete in the labor market for short-term positions compete more often with other workers for a job, and they may therefore display a more competitive attitude. All these factors can lead to substantial economic costs and inefficiencies which may cancel out the benefits of a flexible labor demand.

Since the propensity to cooperate and the other-regarding preferences of permanent workers may vary substantially, depending on the type of firm and the business culture of a company, we selected a benchmark which we expected to be associated with the highest propensity to cooperation and level of pro-sociality. Accordingly, we recruited permanent workers of a co-operative (co-op). [1] The importance of cooperation is one of the distinctive features of a co-op, together with the importance given to the values of self-help, self-responsibility, democracy, equality, equity and solidarity. Hence, one would expect co-op permanent workers to be particularly prone to cooperation (see, e.g., [8]), with significant and sizeable differences with respect to non-co-op workers, in particular temporary workers who are not members of a co-op. *A priori*, this makes the sample of permanent co-op workers a reasonable upper-bound benchmark to be compared with the preferences and behavior of the other two samples. Students, instead, are expected to be the lower bound benchmark for the propensity to cooperate and the other-regarding preferences of temporary and permanent workers, as documented in the experimental literature showing that students are in general less prosocial than non-students (e.g., [6,8,13,14]).

In our experiment, subjects played a Dictator Game, a Decomposed Game and a Public Good Game. These are standard tasks used in the experimental literature to investigate other-regarding preferences and cooperative behavior.

Our results on the behavior of undergraduate students are similar to those reported in the literature with similar subject pools (e.g., [5,8,15–18]). Temporary workers, instead, are significantly more cooperative and less opportunistic. In addition, they are also more other-regarding than students and their behavior and

---

[1] Co-ops are defined by the International Co-operative Alliance as "autonomous associations of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically-controlled enterprise".

preferences are very much similar to those of co-op permanent workers. This result is notable as temporary workers did not know each other, and they shared very little common experiences, either in terms of education, training on the job or specialization. Moreover, the probability of meeting again in subsequent occasions was small, which makes reputation a rather weak argument for inducing cooperation in a repeated interaction perspective. On the contrary, permanent workers knew each other and interacted together on a daily basis in the same workplace. [2] In addition, they worked for the same co-operative, and were constantly exposed to the values of the institution they worked for.

To the best of our knowledge, this is the first paper which, in addition to the comparison between students and workers, also applies the experimental methodology to investigate preferences and propensity to cooperate of temporary and permanent workers. [3] This paper provides complementary evidence to the literature studying workers' effort provision in a real work environment (e.g., [24,25]) and to more traditional sources of information, such as those based on surveys and questionnaires (e.g., [26,27]). In particular, we obtain information on the preferences and propensity to cooperate of co-op workers, on which previous experimental investigation is surprisingly scarce. A noteworthy exception is Burks *et al.* [8], who provide field experimental evidence that employees at firms that pay for performance are significantly less cooperative than those who are members of cooperatives. The comparison between temporary and permanent (co-op) workers is a useful starting point to understand whether these differences are driven by a different propensity to cooperate of workers in general, and whether the possible differences in preferences and behavior may be linked to the specific type of contractual arrangement.

The paper is structured as follows. In next section we describe the three sample pools employed in the experiment. In Section 3, we present the experimental design, and in Section 4, we report the results. Section 5 contains a discussion of the results and Section 6 concludes.

---

[2] We cannot exclude the possibility that, in a given session, some of the students and/or temporary workers knew each other from outside the lab. However, this possibility was minimized by the random process adopted to recruit subjects. In addition, interactions between participants were anonymous, and, therefore, subjects could not know whether they were matched with an acquaintance in a given task.

[3] Other studies investigated the economic and social impact of temporary work arrangements such as, e.g., the existence of a gap in the working conditions of permanent and temporary workers, particularly in terms of working rights, training received, and wage differentials [19], the role of temporary work as a stepping-stone towards regular employment [20], the relationship between temporary work and job satisfaction [21], and the effect of temporary status on the worker's well-being, mental and physical health [12,22,23].

2. The Subject Pools

The subjects taking part in the experiment were recruited from three different populations: undergraduate students of the University of Bologna, temporary employees and permanent employees.

Students. We recruited 96 undergraduate students with an Economics major from the Forlì Campus of the University of Bologna, using the subject pools maintained through ORSEE [28].

Temporary workers. The pool of temporary workers that participated to the experiment was supplied by Obiettivo Lavoro (OL), a large recruitment agency that operates mainly in Italy, with branches operating also in Bolivia, Brazil, Ecuador, Poland, Romania, Paraguay and Peru. OL operates in sectors such as health care and social assistance, cleaning and logistics, construction, large scale retail, hotel, catering, and tourism, with yearly total revenue of 308 million euros (operating revenue in 2012). It manages more than 1,400,000 profiles of workers to be hired to firms or institutions demanding labor force. Originally established as a co-operative company in 1997, OL was converted into a public limited company in 2003. It has adopted an Ethics Code and in 2007 it obtained the SA8000 certification for its Corporate Responsibility practices. [4] The 72 subjects which OL supplied for the experiment were temporary workers who had never participated in an economic experiment like ours, nor ever had relations with OL before 2003.

Permanent workers. The pool of cooperative members that took part in the experiment was supplied by Formula Servizi (FS), a workers co-operative company founded in 1975, which operates in five Italian regions (mainly in Emilia Romagna, where 75% of its revenue is raised). It supplies several services, mainly in sectors such as cleaning, catering, maintenance and logistics, with a yearly total revenue of 55 million euros (operating revenue in 2012). FS has obtained the SA8000 certification for its Corporate Responsibility practices in 2003, and has adopted an Ethical Code in 2012. The total number of workers in 2012 is 1892, with a prevalence of women (83%) and an average age of 47 years. Among these workers, 892 are members of the cooperative ("soci"), 890 are employees, while 110 are outsourced workers. For our experiment, we recruited 84 subjects only among cooperative members, therefore choosing to focus on the most "permanent" set of workers, and we excluded both

[4] SA8000 is a global social accountability standard for decent working conditions. The standard is based on the principles of international human rights and conventions of the International Labour Organisation (ILO), which encourage management to implement sustainable systemic changes in business operations. SA8000 aims to improve the work conditions in organizations; the standard governs child or forced labor, health and safety, freedom of association and right to collective bargaining, discrimination, disciplinary practices, work hours, compensation, and management system.

employees and outsourced staff. None of the permanent workers had participated in economic experiments before.

The details about the socio-demographic characteristics (collected in a questionnaire at the end of the experiment) of the three subject pools are in the Appendix. Students differ in many dimensions compared to both permanent and temporary workers. Most notably, students are younger, less religious, less likely to be married, with a higher education and characterized by a smaller proportion of women. There are also a few differences between permanent and temporary workers. In particular, permanent workers are older, more likely to be married, more likely to be Christian, and with more years of work experience than temporary workers. In all the other dimensions, we do not detect any statistically significant difference between temporary and permanent workers. Because of these differences between our subject pools, we will also control in the data analysis for the socio-demographic characteristics of subjects (especially, those which were significantly different across the three samples) in order to ensure that our results are not due to differences between the characteristics of the three populations. [5]

3. Experimental Design

The experiment is composed of two stages: A classification stage and a main stage. In the classification stage, subjects are required to perform three tasks which are widely employed in the experimental literature to study other-regarding preferences. The subjects first play a one-shot Dictator Game, then they play a Public Good Game to assess their conditional willingness to contribute to a public good (Strategy Method) and finally they play a Decomposed Game. In the main stage, subjects play a repeated anonymous linear Public Good Game (12 repetitions). The experiment was incentivized and fully computerized with the software z-Tree [29]. Anonymity was guaranteed both during the game and the payment procedures.

In the one-shot Dictator Game, each subject is randomly matched with another individual, and has to decide how to divide an endowment of 300 experimental tokens between herself and her matched subject. After each subject has made her decision, a random mechanism establishes whether her proposal or the proposal of the counterpart is implemented. This game provides a measure of fair behavior for each participant. Subjects who give less than 100 tokens are classified as self-centered, while those who give 100 or more are classified as beneficent (for a similar classification, see [17]).

---

[5] Our subjects, and particularly workers, may differ on many other important dimensions (e.g., income, wealth, type of job) for which we do not have information. This limits the possibility to conduct a proper comparison especially between temporary and permanent workers.

In the second task, subjects are classified according to their behavior in a linear one-shot Public Good Game using the Strategy Method Technique [30,31]. Subjects are randomly allocated to groups of four individuals. An endowment of 200 tokens is to be allocated by each subject between a "private" and a "public" account. The individual payoff is determined according to the following function:

$$\pi_i = 200 - g_i + 0.5 \sum_{j=1}^{4} g_j,$$

where g_i is player i's contribution to the public account and g_j the contribution by the other members of the group. Since the contributions of the group to the public account are doubled and equally divided among the group participants, the marginal individual benefit from contributing to the public good is 0.5. Subjects are first asked to make an unconditional contribution to the public good, and then to indicate their willingness to contribute to the public account, conditional upon different possible contributions of the other group members. The possible average contributions range from 0–200, and are listed as multiples of 10 tokens. After the choices are made, one subject per each group is randomly selected and paid according to her conditional contribution to the unconditional contributions of the other three members. The remaining players are paid according to their unconditional contributions. This task provides a measure of subjects' cooperativeness.

In the third classification task, we employ a Decomposed Game Technique, a classificatory task used both in the economic and psychological literature to study the distributional preferences of subjects (e.g., [17,18,31–33]). Participants are asked to choose between two possible allocations of money. For example, a subject must choose whether she prefers an allocation where she receives 130 tokens and the other participant receives 75 tokens, or an allocation where she receives 145 tokens and the other participants loses 39 tokens. Subjects are randomly and anonymously matched in couples and asked to make 24 choices between pairs of allocations. The individual earnings are equal to the sum of the payoffs of the 24 choices made by the subject and by her co-participant, who remains the same throughout the 24 choices (for more details, see [17,18,31]).

No feedback about the co-participants' choices is provided to the subjects during the classification stage: The subjects only receive information about their earnings at the end of the experiment. In addition, the order of the classification tasks is the same for all the subjects, and, therefore, it cannot explain possible across-sample differences.

After the classification stage, the subjects enter the main stage of the experiment and play a repeated linear 4-player Public Good Game (12 rounds). At the beginning of the first round, subjects are randomly matched in groups of four which remain the

same over all rounds. The payoff function is identical to the one used in the Strategy Method, except for the individual endowment (20 tokens instead of 200 tokens). At the end of each round, each subject receives feedback about her contribution to the group account, the single contributions of the other members (whose identities are hidden), [6] and her total earnings. Comprehension and familiarity with the experimental setup is obtained by requiring subjects to enter three forced inputs. [7] After the first 12 rounds, the subjects engage in additional experimental tasks which we do not consider in this paper. [8] After all tasks are completed, subjects are informed about their earnings and are required to complete a demographic questionnaire (reported in the Appendix).

The experiment was conducted at the LES (Laboratorio di Economia Sperimentale, Forlì campus of the University of Bologna, Italy) and at the BLESS (Bologna Laboratory for Experiments in Social Science, Bologna campus of the University of Bologna, Italy) in the period July 2009–January 2011 (details about dates and places of each session are in the Appendix). We ran a total of 18 sessions, with either 12 or 24 subjects per session. Each session was composed by homogeneous subject pools, for a total of 72 temporary workers, 84 permanent workers and 96 students, none of which had previously participated in a similar experiment. No subject could participate in more than one session. The participants were randomly assigned to computer terminals, which were separated by partitions in order to avoid facial or verbal communication between subjects. Before proceeding with each task, subjects filled in a control questionnaire designed to check their understanding of the instructions. Clarifications were individually given to subjects who answered incorrectly. The experiment used a fictional currency, with one token being equal to one euro cent. The exchange rate was not differentiated among the three pools. [9] At

---

[6] The order in which the contributions of the other members is displayed is randomized in each round.

[7] As shown in [34], in a Public Good Game, on-screen instructions requiring forced inputs improve on subjects' comprehension and familiarity with the experimental task. They also contribute to reduce both decision and waiting times without affecting the overall pattern of contributions.

[8] The additional tasks consist of 24 additional rounds of the Public Good Game with some design changes with respect to the initial 12 rounds, and a Stag Hunt Game with framed instructions. Since the subjects were not informed about the content of the future tasks, no distortion in their choices was artificially introduced by the experimenters. To very briefly summarize the findings of these additional tasks, the contributions of permanent and temporary workers remain high all over the 24 additional rounds, whereas the contributions of students drop dramatically after restarting the Public Good Game. This pattern is consistent with the first 12 rounds of the Public Good Game reported in the main text. In the Stag Hunt Game, permanent and temporary workers display very similar behaviors, while students are less likely to cooperate compared to both permanent and temporary workers. More details about these additional tasks are available from the authors upon request.

[9] In the literature, two different approaches have been adopted for paying subjects belonging to different pools. One approach uses incentives which are proportional to the respective outside options, which could imply higher payments for workers and lower ones for students. Another approach (see, e.g., [4]) uses the same incentives and exchange rate, arguing that doing otherwise would introduce

the end of the experiment, subjects were paid for all tasks, on top of a show-up fee of €2. On average, subjects earned €20.93 (approximately 28 US dollars). To secure anonymity in the lab, assistants paid the participants privately at their desks. Each session lasted on average 2 h, including instructions and check questions to verify the understanding of the rules by the participants. The experimental instructions were as neutral as possible, and made available both on screen and on paper at the beginning of each experimental task. To ensure common knowledge, instructions were also read out aloud by the experimenter.

4. Results

In this section, we report the results of the classification tasks (Dictator Game, Public Good Game with Strategy Method and Decomposed Game Technique) and of the repeated linear Public Good Game. [10]

4.1. Dictator Game

Following [17], we classify subjects as beneficent when they donate one third or more of their endowment to the other player, and we classify them as self-centered otherwise. The results are reported in Table 1.

Table 1. Beneficent and self-centered subjects in the Dictator Game.

	Temporary	**Permanent**	**Student**	**Total**
Number of subjects	72	84	96	252
Average offer	40% (121)	45% (135)	23% (70)	35% (106)
Self-centered	25% (n = 18)	14.3% (n = 12)	55.2% (n = 53)	32.9% (n = 83)
Beneficent	75% (n = 54)	85.7% (n = 72)	44.8% (n = 43)	67.1% (n = 169)

We can reject the hypothesis that being beneficent is independent of whether a subject is a temporary worker, a permanent worker or a student (Chi-squared test, $p < 0.001$). [11] The proportion of beneficent subjects among temporary and permanent workers is significantly larger than among students (Chi-squared test, $p < 0.001$). In comparison with the literature, note that the share of beneficent subjects among our undergraduate students is lower than the one found in [17], where the majority of subjects (61%, *i.e.*, 43 out of 71) was classified as beneficent. However, the average offer by students is in line with the literature (see, e.g., [35]). Although beneficent

confounders. The distinction between the two approaches is relevant when the outside options are significantly different. In our experiment, however, the workers' hourly wage is about 10 euros, which is close to the average hourly monetary payoff our laboratory generally pays to students subjects.

[10] In this section we distinguish between permanent and temporary workers. In the Appendix B we show that similar results hold if we pool the two samples together.

[11] Throughout the paper, bivariate tests are two-tailed.

subjects are more common among the permanent workers than among the temporary workers, the difference is only weakly significant (Chi-squared test, $p = 0.091$) and disappears once we control for other covariates (see next paragraph on the regression analysis). [12] This allows stating the following:

Result 1: *In the Dictator Game student are less beneficent than workers. Temporary and permanent workers do not differ.*

Table 2 reports the results from the regression analysis. Regression 1 is a Tobit regression where the dependent variable is the amount offered in the Dictator Game; Regression 2 is a Probit regression on the likelihood of being a beneficent subject. In both regressions the explanatory variables include age and dummies identifying the experimental sample (using temporary workers as baseline category), gender (Male = 1 for male subjects), marital status (Married = 1 for married subjects), religious affiliation (NoReligion = 1 for atheist or agnostic subjects), and educational background (Degree = 1 for subjects with a university degree). [13]

Table 2. Regression analysis (Dictator Game).

	Regression 1			**Regression 2**		
	β	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**
Student	−53.557 ***	12.35	0	−0.181 ***	0.06	0.005
Permanent	3.082	15.77	0.845	0.037	0.09	0.684
Male	4.264	11.3	0.706	−0.020	0.06	0.725
Married	5.092	12.75	0.69	0.028	0.09	0.748
NoReligion	−21.08	12.99	0.106	−0.059	0.07	0.379
Degree	10.108	11.08	0.363	0.042	0.06	0.478
Age	0.945	0.64	0.144	0.009 **	0.00	0.031
Constant	86.641 ***	22.18	0			
Obs	252			252		
Pseudo R2	0.021 [a]			0.141		
Prob > F	0.000			0.000		

Notes: Regression 1: Tobit regression with robust standard errors, the dependent variable is the offer in the Dictator Game; Regression 2: Probit regression with robust standard errors on the likelihood of being a beneficent (in the table, we report the average marginal effects of the independent variables). [a] This is the McFadden's pseudo R-squared. Obs = observations. Prob > F = *p*-value of F-test. dy/dx = marginal effects. ** $p < 0.05$, *** $p < 0.01$.

[12] In the Appendix we show that similar results hold when considering the distribution of offers.

[13] Throughout the paper, in the regression analysis we control for socio-demographic characteristics (see the Appendix). We also ran regressions without including the individual characteristics of the subjects among the explanatory variables. The results qualitatively replicate those reported in the paper and are available upon request.

We detect a significant effect of age in Regression 2. In particular, older subjects are more likely to be beneficent than younger. Once we control for the individual characteristics, the regression analysis is consistent with Result 1.

4.2. Public Good Game with Strategy Method

Using the Strategy Method, subjects are classified in six categories depending on their conditional contributions in the Public Good Game. Similar to [30] we classify as *Conditional Cooperators* those subjects that display a significant monotonically increasing pattern between their conditional contribution and the average contribution of the other group members (Spearman rank correlation $\rho > 0$, $p < 0.01$); as *Free Riders* those who always contribute zero, and as *Triangle* contributors those who display a significant monotonically increasing pattern up to a maximum (Spearman $\rho > 0$, $p < 0.01$) and a significant monotonically decreasing pattern after that maximum (Spearman $\rho < 0$, $p < 0.01$). We also add two categories. Subjects who contribute a positive amount irrespective of the others' contributions are classified as *Unconditional Cooperators*, whereas subjects who display a significant monotonically decreasing pattern (Spearman $\rho < 0$, $p < 0.01$) are classified as *Compensators*, as they seem to counterbalance low contributions by the others. All remaining subjects are pooled in a residual category called *Others*. In Figure 1, we report the average contribution of each type of subject, conditional on the average contribution of the other group members.

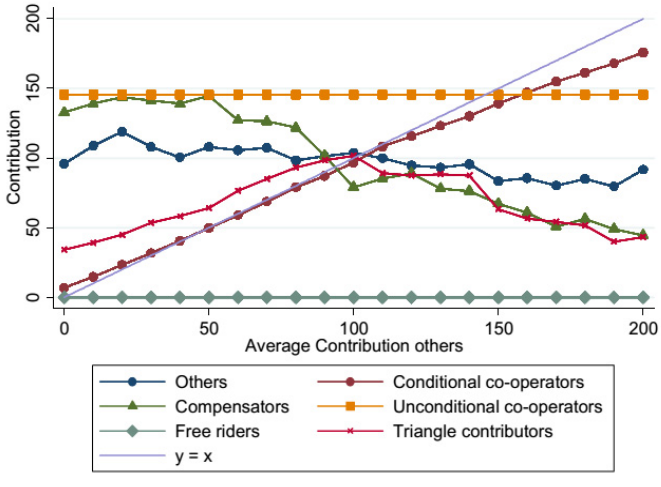


Figure 1. Average contribution (Strategy method), all subjects ($n = 252$).

Table 3 shows the frequencies of the different types of subjects in the three samples and in the whole data set.

We find that the share of conditional cooperators is the largest in all the three samples, which suggests a general tendency to reciprocate and to conform to the choices of the other group members. When considering students, however, some differences deserve to be mentioned. The share of conditional cooperators among students is significantly higher than among permanent workers (Fisher's exact test, $p = 0.029$). Moreover, the share of unconditional cooperators is the lowest among students ($p = 0.040$ when comparing with the permanent workers, and $p = 0.012$ with the temporary workers). [14] The finding that conditional cooperators are the largest share of the sample is consistent with [30], in which they report 50% of the sample ($n = 44$) to be conditionally cooperative, which is admittedly much lower than the results reported in Table 3. In our experiments, the remaining categories do not cover large shares of the population; in contrast Fischbacher *et al.* [30] find that 30% of their subjects behave as a free rider and 14% is a "triangular" subject (called "hump shaped" in their paper). A possible reason for these discrepancies may be due to the different payoff function, and in particular, to a public good multiplier equal to 0.4, instead of 0.5 as in our design.

Table 3. Classification of types in the three samples (Strategy method).

Type	Temporary	Permanent	Student	Total
Conditional cooperator	70.8% ($n = 51$)	64.3% ($n = 54$)	78.1% ($n = 75$)	71.4% ($n = 180$)
Unconditional cooperator	9.7% ($n = 7$)	7.1% ($n = 6$)	1.0% ($n = 1$)	5.6% ($n = 14$)
Free rider	1.4% ($n = 1$)	3.6% ($n = 3$)	5.2% ($n = 5$)	3.6% ($n = 9$)
Triangular	1.4% ($n = 1$)	3.6% ($n = 3$)	2.1% ($n = 2$)	2.4% ($n = 6$)
Compensator	5.6% ($n = 4$)	5.9% ($n = 5$)	2.1% ($n = 2$)	4.4% ($n = 11$)
Others	11.1% ($n = 8$)	15.5% ($n = 13$)	11.5% ($n = 11$)	12.7% ($n = 32$)
Total	100% ($n = 72$)	100% ($n = 84$)	100% ($n = 96$)	100% ($n = 252$)

A priori we expected permanent workers to display a larger share of cooperators (either conditional or unconditional) with respect to temporary workers. Interestingly, this is not the case as the share of cooperators is not statistically different between the two samples of workers ($p > 0.1$).

[14] In the experiment, we also elicited subjects' beliefs about the average contributions by the other group members (see Table A3 in the Appendix). Beliefs do not significantly differ across samples. Interestingly, the average contributions of both permanent and temporary workers are statistically higher than their beliefs about the contribution of the others (Wilcoxon signed-rank test $p < 0.001$ for both), whereas contributions and beliefs of students do not statistically differ (Wilcoxon signed-rank test $p = 0.609$), which is consistent with the evidence that permanent and temporary workers are more likely to be unconditional cooperators compared to students. A Tobit regression where the dependent variable is the unconditional contribution and the set of regressors includes the individual beliefs shows that individual contributions are positively related to the individual beliefs about the contribution of others. The results of the Tobit regression are reported in the Appendix (Table A4).

Table 4. Multinomial logit (Strategy method).

	Compensator			Conditional			Free Rider			Triangular			Unconditional		
	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	−0.021	0.04	0.554	0.091	0.09	0.292	0.010	0.03	0.768	0.005	0.03	0.878	−0.109 *	0.06	0.077
Permanent	−0.022	0.03	0.499	−0.113	0.08	0.165	0.099 **	0.05	0.028	0.028	0.03	0.287	−0.020	0.04	0.636
Male	−0.006	0.02	0.797	−0.051	0.06	0.386	0.015	0.02	0.542	0.020	0.02	0.396	0.010	0.03	0.748
Married	0.065 *	0.03	0.051	0.341 ***	0.12	0.005	−0.388 ***	0.12	0.001	0.021	0.02	0.352	−0.030	0.03	0.357
NoReligion	0.009	0.03	0.772	0.055	0.07	0.460	0.026	0.03	0.335	−0.011	0.02	0.666	−0.007	0.04	0.854
Degree	−0.016	0.04	0.644	−0.031	0.07	0.657	0.010	0.03	0.721	0.033	0.03	0.269	−0.045	0.04	0.264
Age	0.000	0.00	0.902	0.003	0.00	0.509	−0.008 **	0.00	0.019	−0.000	0.00	0.964	0.001	0.00	0.540
Obs								252							
Pseudo R2								0.088							
Prob > F								<0.001							

Note: Multinomial logit regression with robust standard errors. The table reports the average marginal effects of the independent variables for each category of the dependent variable. Obs = observations. Prob > F = p-value of F-test. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Result 2: *The shares of conditional and unconditional cooperators do not differ among temporary and permanent workers.*

We also ran a multinomial logit regression where the dependent variable is the strategy method classification and where we control for the socio-demographic characteristics of the subjects. [15] For certain categories the sample size is very small, and, therefore, the results should be taken with care. In Table 4, we report the average marginal effects of the independent variables for each classification category. Compared to temporary workers, students are less likely to be unconditional co-operators, while permanent workers are more likely to be free riders. Interestingly, married people are more likely to be conditional cooperators and less likely to be free riders. Consistent with the literature, we also detect an effect of age whereby older subjects are less likely to be free-riders, although the effect is small in magnitude.

4.3. Decomposed Game Technique

The Decomposed Game Technique allows classifying subjects in five categories (*Aggressive, Competitive, Individualistic, Reciprocating and Cooperative*) depending on their choices across 24 pairs of allocations. To classify subjects, we consider the total amount of tokens x each person allocated to herself in the 24 choices, and the total amount of tokens y she allocated to her partner. The classification depends on a measure, called the *motivational vector* that is calculated as the inverse tangent of the ratio y/x (see [36,37]). Geometrically, this measure represents the slope of the line passing for the origin and the point (x, y). Subjects are classified as *Aggressive* if the motivational vector has a slope between -112.5 and -67.5 degrees, *Competitive* if it is between -67.5 and -22.5, *Individualistic* if between -22.5 and 22.5, *Reciprocating* if between 22.5 and 67.5, and *Cooperative* if between 67.5 and 112.5. Finally, subjects are classified as *Others* if the length of their vector is less than 75 (see [17,18]). [16]

The results of this classification are reported in Table 5. Comparing the proportion of each type of subject across the three samples, the share of Competitive, Cooperative, Reciprocating and Individualistic subjects statistically differ across the three samples (Fisher's exact test, p = 0.039, 0.038, 0.071, and 0.000, respectively).

Two categories collect more than 75% of the subjects in each sample: Individualistic and Reciprocating. Considering these two categories, the shares are unbalanced, as there are more Reciprocating subjects among permanent workers and more Individualistic subjects among students. Comparing permanent workers

---

[15] We also tried a multinomial probit regression but it does not achieve convergence due to the fact that we have very few cases for certain categories.

[16] The list of choices and the graphical representation of the motivational vector are reported in the Appendix. The terminology used to label the behavioral types is like in [18,34].

and students, the difference in the share of Individualistic subjects is statistically significant (Fisher's exact test, $p < 0.001$), while comparing permanent and temporary workers the share of Individualistic subjects is lower among the former category, although the difference is only weakly significant (Fisher's exact test, $p = 0.088$). When comparing students and temporary workers, the share of Individualistic subjects is higher among students, although the difference is only weakly significant (Fisher's one sided test, $p = 0.080$).

Table 5. Distribution of types in the three samples.

Type	Temporary	Permanent	Student	Total
Aggressive	0% ($n = 0$)	1.2% ($n = 1$)	2.0% ($n = 2$)	1.2% ($n = 3$)
Competitive	4.2% ($n = 3$)	10.7% ($n = 9$)	2.0% ($n = 2$)	5.6% ($n = 14$)
Cooperative	2.8% ($n = 2$)	5.9% ($n = 5$)	0% ($n = 0$)	2.8% ($n = 7$)
Individualistic	41.7% ($n = 30$)	22.6% ($n = 19$)	55.2% ($n = 53$)	40.5% ($n = 102$)
Reciprocating	44.4% ($n = 32$)	54.8% ($n = 46$)	37.5% ($n = 36$)	45.2% ($n = 114$)
Others	6.9% ($n = 5$)	4.8% ($n = 4$)	3.1% ($n = 3$)	4.8% ($n = 12$)
Total	100% ($n = 72$)	100% ($n = 84$)	100% ($n = 96$)	100% ($n = 252$)

Focusing on the students' pool, our results are compatible with those reported in [17], who finds that Individualistic people are the majority (59.9% of her sample) and Reciprocating (which in her taxonomy are called "cooperative") is the second largest category (37.3%).

We can also check whether the information on distributional and motivational characteristics is consistent among the different classification tasks employed in the experiment. Overall, there is a good degree of consistency although correlations are not perfect. For example, being classified as Beneficent in the Dictator Game is positively correlated to being classified as Reciprocating (Spearman $\rho = 0.268$), [17] and negatively related to being classified as Individualistic ($\rho = -0.282$) or as a Free rider ($\rho = -0.318$) in the Decomposed Game. Being classified as a Beneficent in the Dictator Game is also negatively correlated to being a Free Rider ($\rho = -0.138$) in the Strategy Method. Considering each group separately, we observe additional differences. In the permanent workers sample, being classified as Reciprocating in the Decomposed Game is negatively related to being classified as Free rider ($\rho = -0.239$) in the Strategy Method. Interestingly, being classified as an Individualistic in the Decomposed Game is also positively related to being classified as Free rider ($\rho = 0.265$) in the Strategy Method. In the temporary workers sample, instead, correlations between the different classifications are not significant at the 5% level.

[17] We report the correlations that are statistically significant at 5% level. The interpretation of some of these correlations should be taken with caution since, for certain categories, we have very few observations.

We also ran a multinomial logit regression where the dependent variable is the Decomposed Game Technique classification and where we control for the socio-demographic characteristics of the subjects. [18] For certain categories the sample size is very small, and, therefore, the results should be interpreted with caution. Table 6 reports the marginal effects, for each category of subjects, of this regression. The results confirm that workers are more likely to be cooperative compared to students.

4.4. Linear Public Good Game

The classification tasks reveal that the shares of Unconditional cooperators, Conditional cooperators and Free riders differ among the three samples. In a repeated strategic game this allows making some predictions on the behavior of the subjects, since in the presence of conditional cooperators, feedback on the others participants choices and own earnings should induce adaptation to the behavior of other players and, possibly, learning. In particular, since there is a higher share of Unconditional cooperators among the permanent and temporary workers, as well as a lower share of Free riders, we would expect higher initial levels of contributions in the samples of workers than in the sample of students. Furthermore, we would expect Conditional cooperators to provide high levels of contribution in the subsequent rounds. In contrast, since in the student sample there is a higher share of Free riders, we would expect lower initial contributions, which in turn should induce Conditional cooperators to lower their contributions over time.

[18] We also tried a multinomial probit regression but it does not achieve convergence due to the fact that we have very few cases for certain categories.

Table 6. Multinomial logit (Decomposed).

	Aggressive			Competitive			Cooperative			Individualistic			Reciprocating		
	dy/dx	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**
Student	0.185 *	0.11	0.090	−0.034	0.06	0.582	−0.369 ***	0.13	0.006	0.117	0.10	0.243	0.092	0.12	0.437
Permanent	0.195 *	0.12	0.092	0.054	0.05	0.231	0.009	0.02	0.590	−0.246 **	0.11	0.021	0.035	0.10	0.730
Male	−0.182 *	0.11	0.089	−0.045	0.04	0.259	0.035 *	0.02	0.060	0.096	0.08	0.231	0.107	0.08	0.160
Married	−0.184 *	0.11	0.093	0.018	0.04	0.610	−0.023	0.02	0.276	0.074	0.10	0.466	0.103	0.10	0.292
NoReligion	−0.179 *	0.11	0.089	−0.004	0.05	0.948	−0.375 ***	0.13	0.004	0.185 *	0.10	0.055	0.345 ***	0.11	0.002
Degree	0.014	0.01	0.305	0.008	0.03	0.793	0.004	0.02	0.865	0.073	0.07	0.268	−0.062	0.07	0.395
Age	−0.001	0.00	0.297	−0.002	0.00	0.337	0.002 **	0.00	0.044	−0.001	0.00	0.777	0.000	0.00	0.924
Obs								252							
Pseudo R2								0.101							
Prob > F								<0.001							

Note: Multinomial logit regression with robust standard errors. The table reports the average marginal effects of the independent variables for each category of the dependent variable. Obs = observations. Prob > F = *p*-value of F-test. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2 shows the average contribution in the first 12 rounds of the Linear Public Good Game for each subject pool (students, permanent workers, and temporary workers). In Figure 3 we pool workers to highlight the differences between workers and students. Initial contributions do not differ significantly, although we can observe some differences as the game unravels. In the students' sample, average contributions per round and total average contributions are below those by permanent and temporary workers. Comparing the contributions across the three samples for each round, [19] we reject the hypothesis that the three samples come from the same population only for the last two rounds (Kruskal Wallis test, $p = 0.084$ and 0.026, respectively). Students contributed significantly less in the last rounds of the Public Good Game compared to permanent (round 10, 11, 12; Mann-Whitney test, $p = 0.090$, 0.053, and 0.024, respectively) and temporary workers (round 11, 12; $p = 0.069$, and 0.020, respectively), but we do not detect any significant difference in the average contribution of permanent and temporary workers ($p > 0.1$). The students' contributions significantly decrease over time (Spearman $\rho = -0.192$, $p < 0.001$). This is a recurrent finding in the literature (see, e.g., [15,16,35]). Interestingly, there is no significant downward trend for temporary and permanent workers (Spearman $\rho = -0.016$ and -0.060, $p = 0.820$ and 0.343), with an average contribution that remains quite constant throughout the 12 rounds of the Public Good Game. Hence, we can state the following:

Result 3: *Students contribute less than workers and their contributions decrease over time. Temporary and permanent workers contribute on average more than 70% of their endowment and their contributions remain stable over time.*

To study how contributions depend on the history of the game, we employ a Poisson-logit maximum-likelihood hurdle model to separate the decision of contributing in two steps (see, e.g., [38]). First, with a logit model we study whether the subjects decide to contribute or not, then with a zero-truncated Poisson model we study the decision about the amount of the contribution, conditional on contributing a positive amount.

Table 7 displays the results of this estimation. In a first model (Hurdle model 1), the independent variables are the sample dummies (the baseline category is "temporary"), a time variable, the positive and negative deviations of i's contribution from the average group contribution in period $t-1$, the individual i's contributions made in $t-1$ and $t-2$, and, as in previous regressions, the socio-demographic characteristics of the subjects. In a second model (Hurdle model 2), we also include the classification dummies obtained from the classification tasks. In particular, we

[19] We use the average of each Public Good Game group's contribution as the unit of observation.

consider being Individualistic, Competitive or Reciprocating in the Decomposed Game, Beneficent in the Dictator Game, and being Conditional, Unconditional or a Free rider in the Strategy Method.

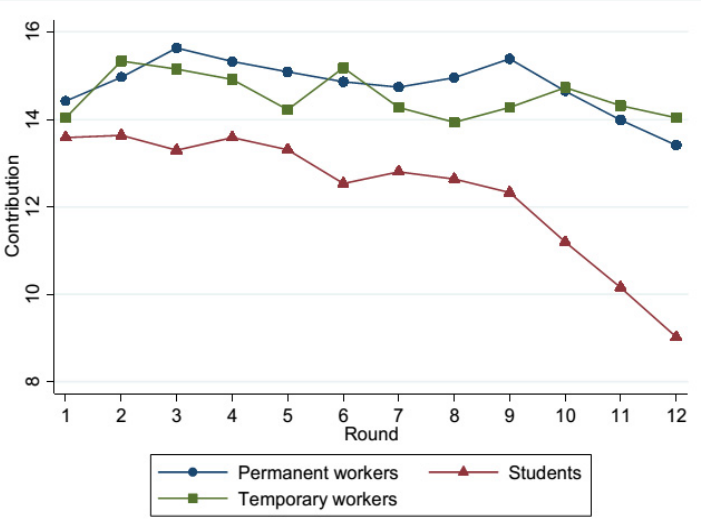


Figure 2. Evolution of average contributions over time (students, permanent, and temporary workers).

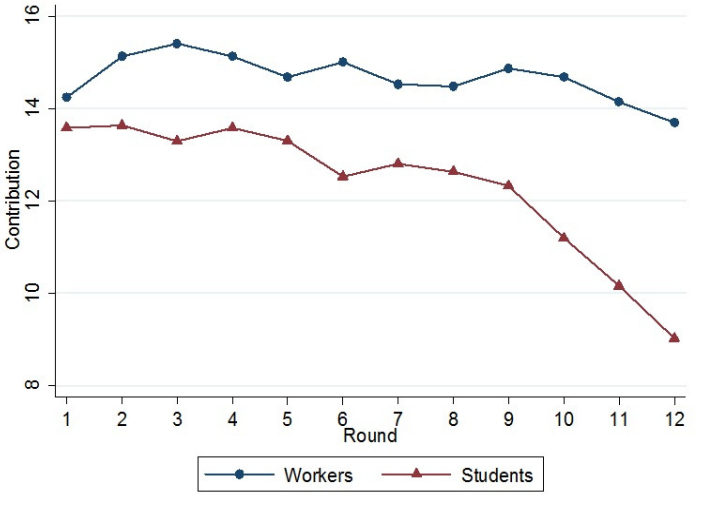


Figure 3. Evolution of average contributions over time (students *vs.* workers).

Table 7. Poisson-logit maximum-likelihood hurdle model.

	Hurdle Model 1					
	Decision to Contribute			**Decision of How Much to Contribute**		
	dy/dx	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**
Permanent	−0.032	0.022	0.134	−0.459 *	0.266	0.085
Student	−0.043 **	0.018	0.016	−0.085	0.264	0.749
Contribution $t-1$	0.007 ***	0.001	0.000	0.738 ***	0.045	0.000
Contribution $t-2$	0.003 ***	0.001	0.002	0.247 ***	0.027	0.000
Positive deviation $t-1$	−0.015 ***	0.002	0.000	−0.305 ***	0.046	0.000
Negative deviation $t-1$	−0.010 ***	0.002	0.000	0.521 ***	0.113	0.000
Period	−0.009 ***	0.002	0.000	−0.030	0.024	0.211
Male	−0.015	0.013	0.278	0.131	0.209	0.531
Married	0.044 *	0.026	0.092	−0.030	0.241	0.899
NoReligion	−0.028 *	0.016	0.075	-0.086	0.256	0.738
Degree	0.022	0.014	0.127	−0.640 ***	0.240	0.008
Age	0.001	0.001	0.235	0.029 **	0.013	0.024
Obs	2520			2292		
	Hurdle Model 2					
	Decision to Contribute			**Decision of How Much to Contribute**		
	dy/dx	**Std. Err.**	***p* > z**	**dy/dx**	**Std. Err.**	***p* > z**
Permanent	−0.033 *	0.019	0.093	−0.369	0.286	0.197
Student	−0.034 *	0.017	0.052	−0.086	0.281	0.758
Contribution $t-1$	0.007 ***	0.001	0.000	0.732 ***	0.045	0.000
Contribution $t-2$	0.002 **	0.001	0.013	0.245 ***	0.027	0.000
Positive deviation $t-1$	−0.015 ***	0.002	0.000	−0.306 ***	0.046	0.000
Negative deviation $t-1$	−0.009 ***	0.002	0.000	0.513 ***	0.113	0.000
Period	−0.009 ***	0.002	0.000	−0.031	0.024	0.202
Male	−0.017	0.012	0.171	0.107	0.214	0.618
Married	0.033	0.024	0.167	−0.051	0.244	0.835
NoReligion	−0.016	0.014	0.274	−0.121	0.258	0.639
Degree	0.008	0.014	0.535	−0.592 **	0.242	0.015
Age	0.001	0.001	0.453	0.028 **	0.014	0.048
Beneficent	0.026 *	0.013	0.054	0.023	0.256	0.929
Reciprocating	−0.110 **	0.051	0.032	0.699	0.472	0.138
Individualistic	−0.115 **	0.051	0.026	0.578	0.485	0.233
Competitive	−0.067	0.060	0.259	−0.165	0.686	0.810
Conditional	−0.005	0.017	0.774	0.062	0.259	0.812
Free rider	−0.091 ***	0.025	0.000	−0.866	1.147	0.450
Unconditional	−0.023	0.022	0.310	0.394	0.491	0.422
Obs	2520			2292		

Notes: The decision to contribute is estimated with a logit model with clustered standard errors at individual level. The decision of how much to contribute is estimated with a zero-truncated Poisson model with clustered standard errors at individual level. In the table, we report the average marginal effects of the independent variables. Obs = observations. Prob > F = p-value of F-test. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The results of the estimation appear to be consistent with our expectations based on the shares of Free riders, Unconditional and Conditional co-operators detected in the classification task (in particular, in the Strategy Method). From the

first model, we observe that students are less likely to contribute a positive amount compared to temporary and permanent workers. This result may be explained by the larger proportion of Free Riders and the smaller proportion of Unconditional contributors among students. Indeed, once we include the classification dummies in the regressions (Hurdle model 2), the point estimate for the variable students decreases and becomes only weakly significant, whereas the variable Free Rider is negative and strongly significant, suggesting that subjects who are classified as Free Riders in the Strategy Method are less likely to contribute to the public good. [20] In both the first and second models, we also find that the probability of not contributing to the public good significantly increases across periods, and it increases if the subject's contribution in the previous period had not matched the average contribution of her group. There is also evidence that subjects are more likely to contribute a positive amount the higher their contribution in the previous rounds (this effect is small, but highly significant). [21] When considering the decision of how much to contribute (conditional on contributing a positive amount), in both models we observe that current contributions are positively correlated to past contributions, and they increase if the subject's contribution in the previous period was smaller than the average contribution of the group. Furthermore, subjects who contributed above the average in the previous period adjust their decision and reduce their contribution in the current round, and older and more educated subjects contribute more than younger subjects. Finally, in the first model we also have some weakly significant evidence that, conditional on contributing a positive amount, permanent workers contribute less than temporary workers.

5. Discussion

In this paper, we investigate two main research questions. First, we study whether we can detect differences in behavior and social preferences between undergraduate students and workers. Second, focusing on the sample of workers, we investigate whether temporary contract and permanent contract workers behave similarly, or whether the type of contractual arrangement is correlated with observable differences.

With respect to the first research question, we find that students tend to be more self-oriented, and less willing to cooperate in a public good game. This result is consistent with similar findings in the literature (for a recent review and a

---

[20] There is also some weakly significant evidence that Reciprocating and Individualistic subjects are less likely to contribute, whereas Beneficent subjects are more likely to contribute.

[21] In the first model, we also find that married subjects are more likely to contribute to the public good, and that atheist or agnostic subjects are less likely to contribute, although these estimates are only weakly significant.

methodological discussion see, e.g., [39]). For instance, using a public good game, Gächter and Herrmann [40] show that students are less prosocial than rural and urban citizens, and Cardenas [41] shows that they extract more resources than rural villagers in a common pool resources experiment. Similar findings are observed when comparing students with workers, such as nurses [5], bicycle messengers [8], shrimp fishermen [13], and white-collar workers [42]. Previous studies suggest that students seem to be at the lower bound of other-regarding preferences. In this paper, we confirm this finding, thereby contributing a little further to the robustness of the results presented in the literature using undergraduate subjects.

Also, we find that there are marked differences between students and workers in the patterns of contributions in the repeated public good game, which suggests that different preferences translate in substantial differences in behavior in strategic settings. The reasons for the observed differences between students and non-students can be attributed to a variety of reasons. For example, in a field study, List [2] suggests that age might play a crucial role, since he finds that younger and middle-aged subjects tend to contribute to a public good at rates consistent with extant laboratory data, whereas older subjects contribute larger amounts of their endowment. Similarly, cooperation rates in prisoner's dilemma games are greater among older than among younger subjects. The existing experiments contrasting students and workers also suggest that the exposition to a working environment might endogenously affect the propensity to cooperate. Which factors lead workers to be more cooperative or to become more opportunistic may in principle depend on several reasons, including the work climate, the type of job, the type of employee or the type of contractual arrangement.

Before running the experiment to address the second research question, we expected to find a significantly lower propensity to cooperate in the group of temporary workers with respect to the group of permanent (co-op) workers, as the former are less likely to know each other and they have received no specific training on the values of cooperation and mutualism. In addition, in terms of social distance co-op permanent workers are closer to each other than temporary workers [43]. The experimental results show, instead, that temporary and permanent workers contribute substantially and in a very similar way to the public good over the 12 rounds of the game. This result holds even if temporary workers tend to be slightly more individualistic than permanent workers.

This finding is suggestive and it deserves a deeper discussion, with the *caveat* that our study has some limitations. In particular, due to the difficulties in recruiting workers as experimental subjects and to the data availability, we cannot provide clean comparisons between the two categories of workers. Taking into account these limitations, we can still make some conjectures on the reasons why the two samples do not differ as much as expected. A possible explanation for the results in the Public

Good Game relies on the large proportion of Unconditional contributors in both the temporary and permanent workers samples (and, analogously, on the low proportion of Free Riders). This may have helped a sustained high level of cooperation for Conditional contributors, thereby maintaining high average contribution rates. In contrast, the higher share of Free Riders among students might have induced Conditional cooperators to lower their contributions over time in response to the low contributions of Free Riders. This interpretation is consistent with the results shown in Table 7, which allows joining and comparing the information on the distributional preferences of the subjects obtained in the Classification stage of the experiment with the Strategic behavior observed in the repeated Public Good Game.

A second possible explanation is that workers, irrespective of whether they are employed under a temporary or permanent contract, are less selfish than students. This explanation would suggest that socialization processes due to working experiences and age could be major factors. If workers have learnt that cooperating provides long-run benefits, they may have brought this experience into the lab and use it when contributing in the Public Good Game. Students, instead, are more used to individual rather than team work and have less experience with repeated interactions and the related reputational concerns. Moreover, students participating in lab experiments may be a selected population with a stronger focus on earning money [31]. Unfortunately, our design does not allow disentangling this kind of selection issues. [22]

Subjects only knew that the pools were homogeneous. This might have induced an in-group bias. For example, temporary workers may have chosen their contributions based on mutual insurance considerations. Accordingly, they might have contributed to the public good knowing that the other group members were also temporary workers, and this may have stimulated high contributions. However, this effect should be stronger among the permanent workers, as they work for the same cooperative and meet each other daily, but we detect no significant difference with respect to temporary workers. This result shows that temporary and permanent arrangements do not affect the propensity to cooperate to a public good when workers are part of a homogeneous group. As far as we know, this is the first experimental evidence showing this finding. A different, but related, question which we leave for future investigation is whether heterogeneous groups

---

[22] Note that the above results cannot be due to demand effects nor to reputation effects, as we ensured anonymity and, if anything, we should expect this effect to operate in the same direction in all sessions. Similarly, it is unlikely that the results are due to a lack of understanding of the subjects or more systematic mistakes by temporary and permanent workers, since great care was taken to ensure subjects' understanding of the instructions (for example, we included comprehension questionnaires at the end of the instructions, and we allowed plenty of time for individual clarifications).

composed by both temporary and permanent workers display the same levels and trends of contributions.

Our *a priori* expectation of a higher contribution rate among permanent workers was based on the fact that they know each other, that they work in the same place, and that the common employer is a co-op which actively emphasizes the values of mutualism and cooperation. It is however possible that some other factors may have driven the behavior of permanent workers toward an unexpected direction. Knowing each other is not a guarantee *per se* of a high social capital and willingness to cooperate, as human relationships can improve, but as well deteriorate over time. Analogously, being constantly exposed to the values of mutualism does not ensure these values are internalized by workers, since these values may possibly be rejected. If this were the case, we should observe low contribution rates. The result that the observed contribution rates of permanent workers are significantly higher than those of students and that they remain quite high over time (about 60% of the initial endowment), however, does not support this hypothesis.

The high degree of cooperative and other-regarding behavior observed among temporary workers may be due to positive selection and signaling effects. As observed by Engellandt and Riphahn [24], temporarily employed workers may have below average risk-aversion and accept temporary jobs to signal their ability and qualify for permanent positions. In a similar fashion, social and labor psychologists refer to *impression management*, as the systematic attempt by temporary workers to behave in a way that pleases the employee [44]. Such an attitude might have been brought by the temporary workers into the lab, and push them to cooperate. Although this possibility cannot be excluded *a priori*, we made very clear that the participation to our experiment was a one-shot experience, and that in no way we would have communicated the results of their individual choices to the recruitment agency.

It is possible that the common social environment in which both temporary and permanent workers live has played a major role for working people. The social environment can overwhelm the role of investments in the values of cooperation by the co-operative, or the possible differences induced by permanent and temporary work arrangements. This can be the case, as the places where the experiments were conducted (located within the Emilia-Romagna region, Italy) are characterized by high levels of social capital and generalized trust [42]. This may have shaped in a similar way the behavior of both temporary and permanent workers, as well as that of the students raised in Emilia-Romagna. However, this conjecture is not consistent with the finding that students display significantly lower contribution rates even when controlling for age. Hence, if the reason for the scant difference in the contribution rates of temporary and permanent workers relies on the existence of a high social capital at the regional level, one should also explain why this

seems to affect more workers than students. This requires additional data aimed at understanding whether our results also hold in regions where social capital is low. [23]

6. Conclusions

Assessing the robustness of results obtained in a controlled environment using different subject pools is key for the evaluation of the external validity of experimental studies. We address this issue by running an artefactual field experiment using three different subject pools: undergraduate students, temporary contract workers, and permanent contract workers.

The results obtained when contrasting students with workers are consistent with the findings available in the literature: students tend to be less prosocial than workers. We find, instead, only minor differences in the distributional preferences of the two groups of workers, and a very similar behavior in a Public Good Game. Workers, irrespective of the type of their contractual arrangement, begin with high levels of contribution to the public good, and contributions remain high over time. In the sample of students, instead, contributions begin at high levels, but then decline considerably over time.

While the latter result is in line with the previous experimental evidence on the Public Good game, the high and steady level of contributions by workers is a less common finding. A possible explanation for this result is that the different contractual arrangements, and the consequent economic and psychological effects, play a minor role with respect to other factors. For example, socialization and learning on the job may have induced both temporary and permanent workers to behave similarly and avoid free-riding opportunities. Also, in-group and mutual insurance considerations may have driven workers to contribute to the public good. In either case, our paper documents that, in our sample of subjects, students cooperate less than workers, and that temporary and permanent workers cooperate at similar rates. This is a novel and somehow unexpected result which deserves further investigation.

---

[23] Another possible explanation is that, contrary to our priors, co-op workers are not the only workers that had been exposed to the values of cooperation and mutualism. The recruiting agency which supplied the temporary workers was originally established as a co-op and, although it converted into a public limited company in 2003, it adheres to the Social Responsibility practices, it has an Ethical code and its guiding principles are (according to its Charter of Values) attentiveness, collaboration, improvement, equity, transparency, and integrity. We cannot exclude that these factors have had an influence on the preferences and behaviour of the temporary workers that took part in our experiment, but we do not think it has played a major role because there is very little interaction between the temporary workers and the staff of the recruiting agency, which is a simple intermediary between employers and workers. If anything, there may have been an interaction with the temporary employers, some of which also have Ethical codes and adhere to the Social Responsibility practices, while some others do not.

Acknowledgments: We thank Formula Servizi and Obiettivo Lavoro for their precious assistance in recruiting the experimental subjects. We thank Maria Bigoni, Stefania Bortolotti, Marco Casari, Caterina Giannetti, Vera Negri Zamagni, the participants at the 2011 BEELab conference in Florence (Italy), two anonymous referees, and the guest editor Ananish Chaudhuri for comments and suggestions. Financial support from the PRIN 2007/B8SC7A_002, the University of Bologna Strategic Project "Ethical Values and Competitiveness of Italian Cooperative Companies", CFICEI (Centro di Formazione e Iniziativa sulla Cooperazione e l'Etica d'Impresa) and AICCON is gratefully acknowledged. The usual disclaimer applies.

Author Contributions: All the authors contributed equally to this article.

Appendix
A. Participants, Sessions, and Additional Analysis

A.1. Socio-Demographic Characteristics of the Participants

Table A1. Subjects' socio-demographic characteristics.

Characteristics	Permanent	Student	Temporary
Gender	(*n* = 84)	(*n* = 96)	(*n* = 72)
Female	58 (69.05%)	52 (54.17%)	43 (59.72%)
Male	26 (30.95%)	44 (45.83%)	29 (40.28%)
Age	(*n* = 84)	(*n* = 96)	(*n* = 72)
Mean	41.64	23.94	31.67
St. dev.	9.19	4.57	9.39
Marital Status	(*n* = 84)	(*n* = 96)	(*n* = 72)
Married	44 (52.38%)	1 (1.04%)	13 (18.06%)
Unmarried	40 (47.62%)	95 (98.96%)	59 (81.94%)
Nationality	(*n* = 84)	(*n* = 96)	(*n* = 72)
non-Italian [a]	7 (8.33%)	5 (5.21%)	8 (11.11%)
Italian	77 (91.67%)	91 (94.79%)	64 (88.89%)
Religion	(*n* = 84)	(*n* = 96)	(*n* = 72)
Agnostic/Atheist	9 (10.71%)	28 (29.17%)	15 (20.83%)
Christian	71 (84.52%)	64 (66.67%)	42 (58.33%)
Other [b]	4 (4.76%)	4 (4.17%)	15 (20.83%)
Education	(*n* = 84)	(*n* = 96)	(*n* = 72)
Lower Secondary School or less	56 (66.67%)	95 (98.96%)	52 (72.22%)
Upper Secondary School or more	27 (32.14%)	0 (0%)	14 (19.44%)
Other	1 (1.19%)	1 (1.04%)	6 (8.33%)
Years of work	(*n* = 36) [c]		(*n* = 72)
Mean	19.42		10.22
St. dev.	10.59		9.01
Years of work in cooperative	(*n* = 36) [c]		
Mean	10.01		
St. dev.	6.88		
Years of work as temporary			(*n* = 72)
<1			41 (56.94%)
>1			31 (43.06%)
Working days as temporary worker (last 12 months)			(*n* = 72)
Mean			46.04
St. dev.			87.55
Working days as temporary worker (total)			(*n* = 72)
Mean			92.22
St. dev.			189.96
Months worked in cooperative			(*n* = 71) [d]
Mean			14.2
St. dev.			83.97

Table A1. *Cont.*

Characteristics	Permanent	Student	Temporary
Student status			(n = 72)
Only worker			58 (80.56%)
Lower Secondary School or less			1 (1.39%)
Upper Secondary School or more			11 (15.28%)
Other			2 (2.78%)

Notes: [a] Seven were Albanian, 1 Brazilian, 2 Moldavian, 1 Polish, 5 Romanian, 2 Senegalese, 1 Serb, and 1 Hungarian. [b] Two were Buddhist, 1 Indu, 6 Muslim, and 14 did not specify. [c] This information was collected only for a subset of permanent workers. [d] One subject did not specify how many months she worked in a cooperative.

If we compare the proportion of males and females across the three samples, we do not detect any statistically significant difference (Chi-squared test, $p = 0.122$). In pairwise comparisons, the only statistically significant different occurs between permanent workers and students. In particular, the proportion of females is significantly lager in the permanent workers sample compared to the students sample ($p = 0.041$). There is no statistically significant difference in the proportion of females between temporary and permanent workers, and temporary workers and students ($p = 0.224$ and 0.472 respectively). The three samples statistically significantly differ with respect to age (Kruskal Wallis test, $p < 0.001$). In particular, permanent workers are significantly older than students and temporary workers (Mann-Whitney $p < 0.001$ for both, pairwise comparison), and temporary workers are significantly older than students ($p < 0.001$). If we look at the marital status of the subjects, the three samples statistically significantly differ both in aggregate (Chi-squared test, $p < 0.001$) and in pairwise comparisons ($p < 0.001$). In particular, the proportion of permanent workers who are married is significantly larger than students and temporary workers. Similarly, the proportion of married subjects in the temporary workers sample is significantly larger than in the students' sample. Almost the totality of the participants is Italian (92%). Only few subjects are not from Italy. Those who are not Italian are Albanian, Brazilian, Moldavian, Polish, Romanian, Senegalese, Serb, or Hungarian. If we compare the proportion of non-Italian across the three samples, we do not detect any statistically significant difference (Chi-squared test, $p = 0.370$). A similar result is obtained in pairwise comparisons ($p > 0.1$). If we look at the religious affiliation, subjects statistically significantly differ in their religious beliefs across samples (Chi-squared test, $p< 0.001$). This evidence is also supported in pairwise comparisons (Chi-squared test, $p < 0.01$). More specifically, the fraction of Christians is larger among permanent workers than temporary workers and students. Students are generally more agnostic/atheist. The proportion of participants who are neither Christian nor agnostic/atheist is larger among temporary workers compared to students and permanent workers. In general, permanent workers are more religious than students ($p = 0.002$) and temporary workers ($p = 0.081$). In contrast, temporary

workers are not statistically significantly more religious than students ($p = 0.221$). If we compare the level of education, the three samples statistically significantly differ (Chi-squared test, $p < 0.001$). In particular, a higher proportion of students possess a higher degree compared to permanent and temporary workers (Chi-squared test, $p < 0.001$). No statistically significant difference occurs between temporary and permanent workers on the level of education (Chi-squared test, $p = 0.124$). If we compare the years of work of permanent and temporary workers, permanent workers have more years than temporary workers (Mann-Whitney $p < 0.001$).

A.2. Details of the Sessions

Table A2. Details of the sessions.

Session	Sample Pool	Number of Participants	Date (Day/Month/Year)	Location
1	Students	12	29/07/2009	Forli
2	Students	12	29/07/2009	Forli
3	Permanent workers	12	03/09/2009	Forli
4	Permanent workers	12	03/09/2009	Forli
5	Permanent workers	12	04/09/2009	Forli
6	Students	24	14/04/2010	Forli
7	Permanent workers	12	14/04/2010	Forli
8	Permanent workers	12	16/04/2010	Forli
9	Students	24	16/04/2010	Forli
10	Permanent workers	12	16/04/2010	Forli
11	Permanent workers	12	21/04/2010	Forli
12	Students	12	24/02/2011	Forli
13	Students	12	24/02/2011	Forli
14	Temporary workers	12	24/02/2011	Bologna
15	Temporary workers	12	25/02/2011	Forli
16	Temporary workers	12	25/02/2011	Forli
17	Temporary workers	12	04/03/2011	Forli
18	Temporary workers	12	25/03/2011	Bologna
19	Temporary workers	12	21/06/2011	Bologna

A.3. Distribution of Offers in the Dictator Game

Figure A1 shows the distribution of the offers per sample. Comparing the distributions of the offers between permanent workers and students, we reject the null hypothesis that they are the same (Epps-Singleton test, $p < 0.001$). A similar result holds when comparing temporary workers and students (Epp-Singleton test, $p < 0.001$). When we compare the sample distributions of the offers between temporary and permanent workers, we only weakly reject the hypothesis that they have been drawn from the same population (Epps-Singleton $p = 0.077$).

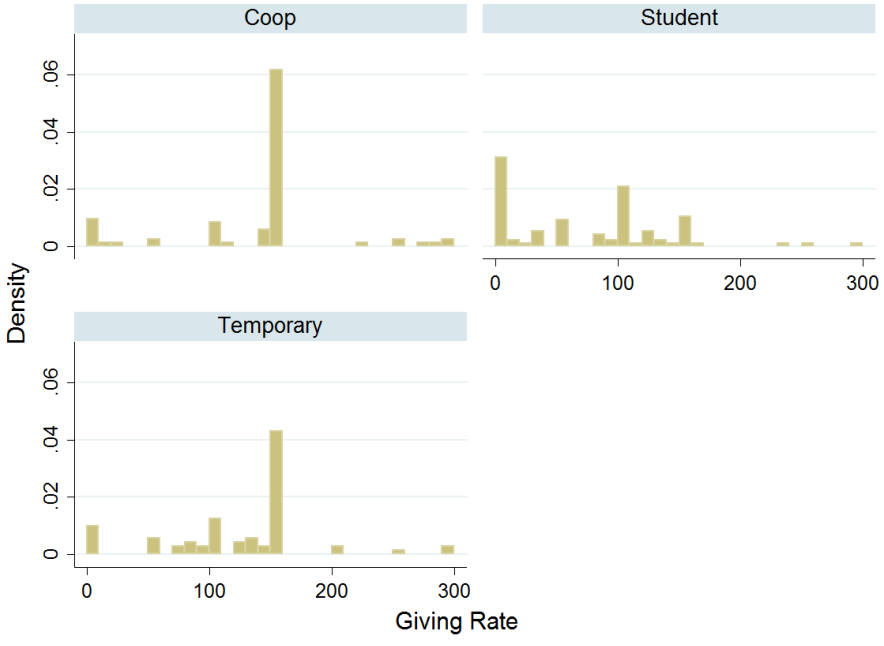


Figure A1. Dictator Game: distribution of offers.

A.4. Unconditional Contributions and Beliefs in the Strategy Method

If we look at the unconditional contributions in the Strategy Method (Table A3), they weakly statistically differ across samples (Kruskal Wallis test, $p = 0.061$). In particular, if we conduct pairwise comparisons, students contributed on average less than permanent and temporary workers. However, the difference is statistically significant only when comparing students with permanent workers (Mann-Whitney test, $p = 0.021$). The beliefs do not significantly differ across samples, or in pairwise comparisons. Interestingly, the average contributions of both permanent and temporary workers are statistically higher than their beliefs about the contribution of the others (Wilcoxon signed-rank test, $p < 0.001$ and 0.001, respectively), whereas contributions and beliefs of students do not statistically differ (Wilcoxon signed-rank test $p = 0.609$). This result support the evidence that permanent and temporary workers are more likely to be unconditional co-operators compared to students, *i.e.*, they contribute no matter what is the contribution of the others.

Table A3. Unconditional contributions and beliefs.

Sample	Contribution		Belief	
	Mean	**Std. Err.**	**Mean**	**Std. Err.**
Permanent (n = 84)	118.99	57.54	102.34	54.05
Student (n = 96)	97.88	60.18	105.71	47.10
Temporary (n = 72)	112.13	53.75	101.19	46.27
Total (n = 252)	108.99	58.02	103.30	49.14

We also ran a Tobit regression where the dependent variable is the unconditional contribution. Independent variables include dummy variables for the experimental sample (using the temporary workers as baseline category), age, gender (Male = 1 for male subjects), relationship status (Married = 1 for married subjects), religious affiliation (NoReligion = 1 for atheist or agnostic subjects), educational background (Degree = 1 for subjects with a university degree), and beliefs. Table A4 displays the results of the regression. When covariates are controlled for, we observe that students contribute significantly less than temporary workers, whereas the unconditional contributions of permanent workers do no statistically differ from those of temporary workers. In addition, there is strong evidence that the contribution is positively related to the beliefs about the contribution of the others. This result does not surprise since about 70% of the subjects were classified as conditional co-operators.

Table A4. Tobit regression (unconditional choice).

	β	**Std. Err.**	$p > z$
Belief	0.991 ***	0.08	0.000
Student	−20.937 **	8.71	0.017
Permanent	6.864	9.05	0.449
Male	−4.767	7.41	0.521
Married	−12.099	9.11	0.185
NoReligion	−9.649	9.24	0.298
Degree	8.562	7.95	0.282
Age	0.574	0.38	0.129
Constant	2.708	15.12	0.858
Obs	252		
ll	−1113.4		
Prob > F	0		

Note: Tobit regression with robust standard errors. Obs = observations. Prob > F = p-value of F-test. ll = log-likelihood. ** $p < 0.05$, *** $p < 0.01$.

Table A5. Choice of allocations.

Question	Option A		Option B	
	Self	Other	Self	Other
1	150	0	145	39
2	144	−39	130	−75
3	130	−45	106	−106
4	106	−106	75	−130
5	75	−130	39	−145
6	39	−145	0	−150
7	0	−150	−39	−145
8	−39	−145	−75	−130
9	−75	−130	−106	−106
10	−106	−106	−130	−75
11	−130	−75	−145	−39
12	−145	−39	−150	0
13	−150	0	−145	39
14	−145	39	−130	75
15	−130	75	−106	106
16	−106	106	−75	130
17	−75	130	−39	145
18	−39	145	0	50
19	0	150	39	145
20	39	145	75	130
21	75	130	106	106
22	106	106	130	75
23	130	75	145	39
24	145	39	150	0

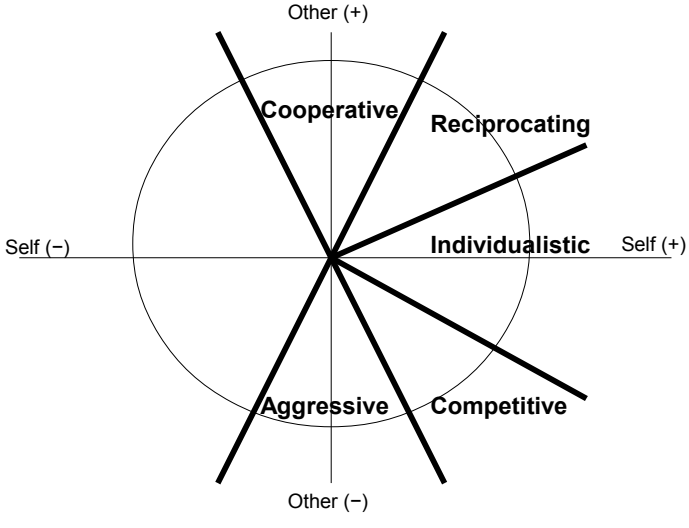


Figure A2. The "value orientation circle".

Introduction

Welcome to the Laboratory of Experimental Economics of the University of Bologna.

You will participate in a study on individual behavior of about one hour and a half. If you read these instructions carefully, you can, depending on your decisions, earn some money. For your convenience, instructions are provided both on screen and on paper.

Your earnings will be calculated in *florins* and will be converted into Euros at the end of today's session. Every florin equals 1 euro cent. Payment will be made in cash at the end of this session and it will be done in such a way that no other participant will know how much you have earned.

The experiment is divided into various stages. At each stage you will be asked to make some decisions or to answer a few simple questions.

From now on it is forbidden to talk to the other participants, or communicate in any other way. If you want to ask a question, raise your hand.

STAGE ONE

Situation

In this stage each of you is randomly matched with another participant. The identity of the participant with whom you are matched remains anonymous. You will never know with whom you are matched.

For each pair of participants, the computer assigns 300 florins, which are randomly given to only one of the two members of the couple. The other participant gets nothing.

Those who receive the 300 florins can decide whether to keep them all or give part of the florins to the other participant they are matched with. Those who have not received the 300 florins, instead, can only receive florins from their partner.

You cannot currently know whether you are one of those who will receive the 300 florins, because those who will receive 300 florins will be randomly selected at the end of stage seven.

What you should do in stage one

At this stage we ask you to indicate how you would divide 300 florins between you and the other person, in case you are randomly selected to receive the 300 florins. If, at the end of stage seven, you actually receive the 300 florins, then your choice will be implemented. If, instead, your partner will be selected, her/his choice will be implemented, and you will receive the amount of florins indicated by her/him.

Push the button "Continue" and make your choice.

STAGE TWO

Situation

In this stage, each of you is randomly matched with three other participants to form a group of four persons. The identity of the other participants is anonymous and you will not know with whom you are matched.

Each person receives 200 florins and must decide how many florins to put in her/his own personal account and how many to invest in a project. The overall payoff is given by the personal account plus the earnings resulting from the project.

Earnings from the personal account: For each florin you put in your personal account you will earn exactly one florin.

For example, if you put 200 florins in your personal account (and therefore you do not invest in the project), you will earn exactly 200 florins. If you put 60 florins in your personal account, you will earn 60 florins.

Earnings from the project: The earnings resulting from the project depend on your choices and the choices of the other members of the group. For each member of the group, the earnings from the project will be determined as follows:

1) All florins given by the group members are summed up.
2) The sum is doubled.
3) The doubled sum is divided into four equal parts and assigned to each group member.

Now let's see two examples to better understand how the earnings from the project are calculated:

Example 1: If the sum of all contributions to the project is 300 florins, each group member will receive individually:

$$(300 \text{ florins multiplied by } 2 \text{ and then divided by } 4) = 150 \text{ florins from the project.}$$

Example 2: If the four members of the group invest 25 florins each, and then the sum of their investment is 100 florins, each group member will receive:

$$(100 \text{ florins multiplied by } 2 \text{ and then divided by } 4) = 50 \text{ florins from the project.}$$

Practice in stage two

Please answer the following questions. The purpose is to practice with the computation of the earnings you will get. The answers you give to these questions

will not affect your final earnings. (If you like, you can use the electronic calculator that you can activate by pressing the small button at the bottom of the screen).

1. Each member of the group has 200 florins at his disposal. Suppose that none of the four members of the group (including you) contributes to the project.
   - How much will your total earnings be (income + personal project)?
   - What will the individual earnings of the others members of the group be?
2. Each member of the group has 200 florins at his disposal. Suppose you put 200 florins in the project and each of the other members of the group puts 200 florins.
   - How much will your total earnings be (income + personal project)?

Start of Stage Two

Now the choice situation we have just described begins. You will be given 200 florins and you will decide how much you want to put in your personal account and how much in the project. The mechanism for calculating earnings is the one just described.

What you should do in stage two

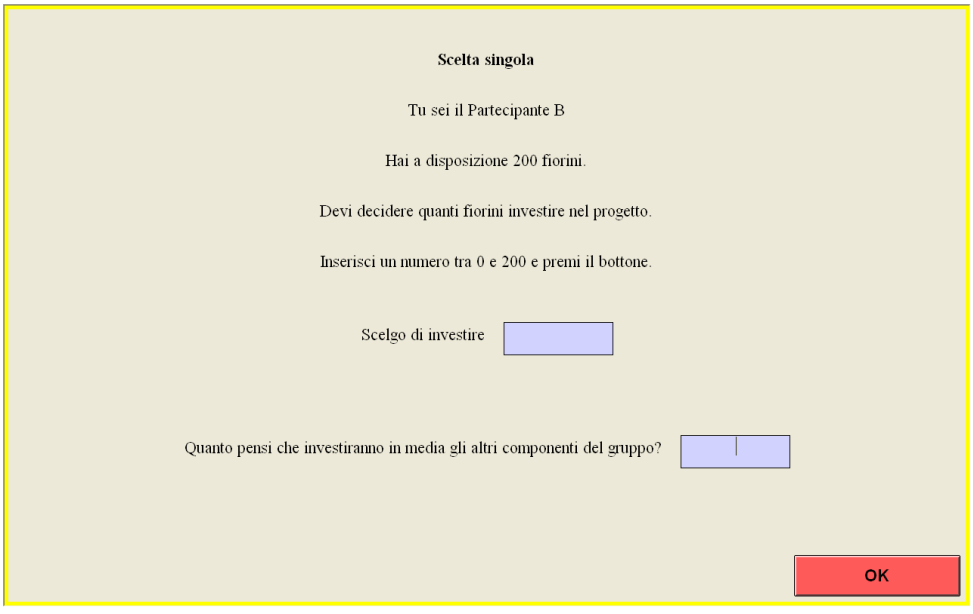
In this stage you will make two kinds of choices: we will call the first "single choice" and the second "choice in the table". Your earning in this stage depends both on what you have chosen in the "single choice", and on what you have chosen in the "choice in the table".

Single Choice

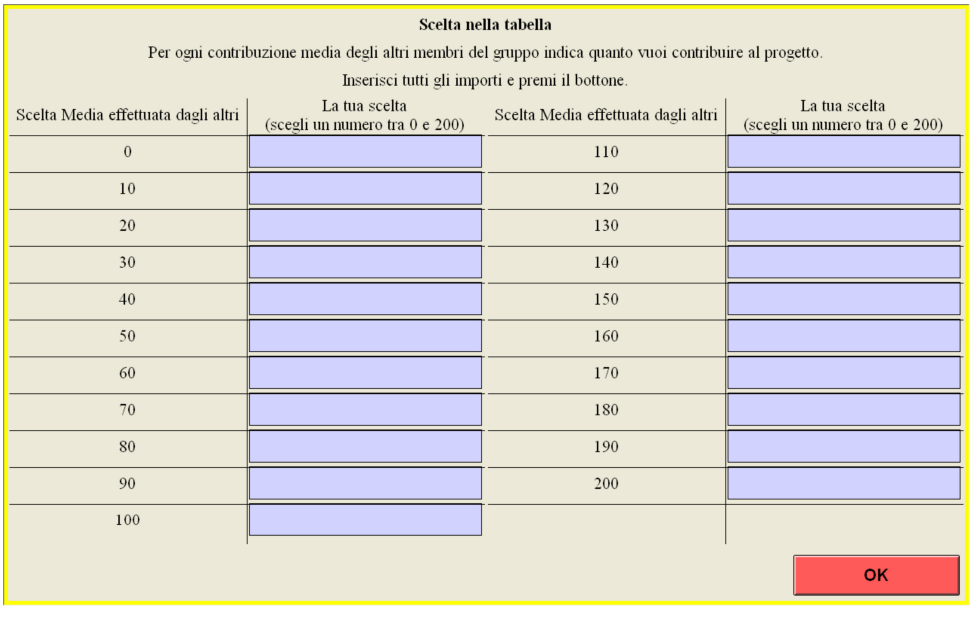
With the single choice you have to decide how many of the 200 florins you want to put in the project.

You also have to indicate how much you think that others are investing in the project. If your guess is at 3 florins or closer from the actual average, you earn 3 extra florins.

On your screen you will see this:



After pressing the "OK" button, you will go to the "choice in the table".
Choice in the table: The screen for the "choice in the table" will look like this:



For example, in the third cell of the first column you are asked to indicate how much you would like to contribute to the project if the average contribution of each of other members of your group is 20 florins. Or, in the third cell of the second column you are asked to indicate how much you would like to contribute if each of the other group members contributes (on average) 130 florins.

Results of stage two

After all participants to the experiment have given their answers, the computer will randomly select a person in each group.

The computer will take the choices made by the other 3 people in the "single choice" and it will compute the average of their contributions. Then it will consider what is the contribution choice indicated by the fourth person in the "choice in the table" in correspondence of the average contribution of the other three members.

By combining this information, the computer:

- Will compute the total contribution of the group by adding the three individual contributions (single choice) and the contribution given by the selected person (choice in the table);
- Will determine the earnings from the project, by doubling the total amount of the contributions to the project;
- Will give a quarter of the doubled sum to each member of the group.

Since you do not know who will be selected by the computer to determine the choice in the table, when you have to fill in the single-choice and the choice in the table, you have to think carefully about both types of choice because both can prove to be decisive in the determination of your earnings.

The computer draw and the result of these computations will be communicated at the end of stage seven. Press the "Continue" button to begin.

STAGE THREE

Situation

At this stage you have to choose between two options, Option A and Option B. The two options are related to sums of money that you and another participant will earn. For example, you may be asked to choose between two options, A and B, where Option A is a gain of 145 florins for you and a loss of 39 florins for another participant, while Option B means a gain of 130 florins for you and a gain of 75 for the other

participant. The other participant will have to choose from the same options. An example of the choice between A and B is the following:

There is a total of 24 pairs of choices, and you will be matched with the same participant (selected at random by the computer at the beginning of stage three) for all 24 pairs of choice. The identity of the other participant will remain anonymous, and no one will know who has been matched with.

Your payoff depends on your choices and on those of the other participant, and it is given by the sum of the choices made by you and the other participant.

Going on with the previous example, if you choose Option A (a gain of 145 florins for you, a loss of 39 loss for your partner), and the other chooses Option B (a gain of 130 florins for her/himself and a gain of 75 florins for you), you will earn 145 + 75 = 220 florins, while the other participant will earn $-39 + 130 = 91$ florins. If, instead, also the other chooses Option A, you will earn 145 + 75 = 220 florins, while the other participant will earn $-39 + 130 = 91$ florins. If, instead, the other also choses Option A, you will earn $145 - 39 = 106$.

What you should do in stage three

In this stage we ask you to make 24 choices between Options A and B. The options will be different each time.

Results

The overall outcome of your choices will be known only at the end of stage seven. In any case, you will be paid privately and you will not know the identity of the other participant, nor she/he will know yours.

Before starting with stage three, we ask you to answer some questions. The purpose is to practice with the computations of the earnings you will get. The answers you give to these questions will not affect your final earnings.

In the situation described below:

If both choose Option B

How much do you earn?
How much does the other participant earn?

If you choose Option B, while the other participant chooses A

How much do you earn?
How much does the other participant earn?

STAGE FOUR

Situation

You will make choices similar to those of stage two, the one in which you received a sum of money and had to choose how much to invest in a common project and how much to put in your personal account.

As before, the sum of money invested in the project is doubled and then divided among four anonymous participants, and your earnings will be the sum of these florins and of those that you put in your personal account.

The participants in your group will be randomly selected by the computer from all the participants of this experiment and they will remain the same throughout this stage.

Unlike before, every time you and the other three participants choose how much to invest in the project, the individual contributions will be immediately communicated to all participants in the group. To protect anonymity, the participants are indicated with letters A, B, C or D that the computer randomly assigns at each round. This means, for example, that letter A can indicate different people in the course of the stage.

Along with your choice of contributing in the project, we will ask you what you think the others are contributing to the project, on average. If your prediction is at 3 or less florins from the actual average, you will earn three extra florins.

What you should do in stage four

This stage consists of 12 rounds. In each round you are given 20 florins and you are asked to choose how much to invest in the project (12 choices), and to indicate what you think is the average contribution of the other participants (12 forecasts).

Results

After you have made your choice and your prediction, you will be informed about the others' contributions and how much you have earned. Payment will take place privately at the end of stage seven.

Table A6. Beneficent and Self-centered subjects in the Dictator Game.

	Worker	**Student**	**Total**
Number of subjects	156	96	252
Average offer	43% (128)	23% (70)	35% (106)
Self-centered	19.2% (n = 30)	55.2% (n = 53)	32.9% (n = 83)
Beneficent	80.8% (n = 126)	44.8% (n = 43)	67.1% (n = 169)

Table A7. Regression analysis (Dictator Game).

	Regression 1			Regression 2		
	β	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	−54.220 ***	12.64	0.000	−0.189 ***	0.06	0.003
Male	4.039	11.11	0.717	−0.022	0.06	0.694
Married	5.701	12.46	0.648	0.036	0.08	0.672
NoReligion	−21.204	12.91	0.102	−0.059	0.07	0.374
Degree	9.822	10.89	0.368	0.039	0.06	0.502
Age	0.996 *	0.52	0.058	0.009 ***	0.00	0.010
Constant	86.345 ***	21.81	0.000			
Obs	252			252		
Pseudo R2	0.021 [a]			0.141		
Prob > F	0.000			0.000		

Notes: Regression 1: Tobit regression with robust standard errors, the dependent variable is the offer in the Dictator Game; Regression 2: Probit regression with robust standard errors on the likelihood of being a beneficent (in the table, we report the average marginal effects of the independent variables). [a] This is the McFadden's pseudo R-squared. Obs = observations. Prob > F = p-value of F-test. dy/dx = marginal effects. * $p < 0.1$, *** $p < 0.01$.

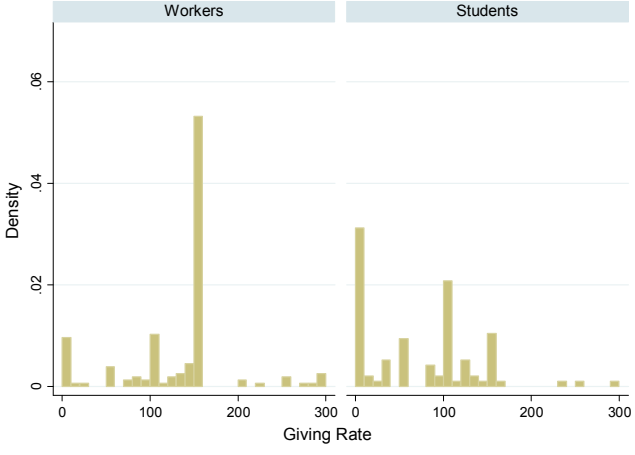


Figure A3. Dictator Game: Distribution of offers.

Table A8. Classification of types in the three samples (Strategy method).

Type	Worker	Student	Total
Conditional cooperator	67.31% (n = 105)	78.1% (n = 75)	71.4% (n = 180)
Unconditional cooperator	8.33% (n = 13)	1.0% (n = 1)	5.6% (n = 14)
Free rider	2.56% (n = 4)	5.2% (n = 5)	3.6% (n = 9)
Triangular	2.56% (n = 4)	2.1% (n = 2)	2.4% (n = 6)
Compensator	5.77% (n = 9)	2.1% (n = 2)	4.4% (n = 11)
Others	13.46% (n = 21)	11.5% (n = 11)	12.7% (n = 32)
Total	100% (n = 156)	100% (n = 96)	100% (n = 252)

Table A9. Multinomial logit (Strategy method).

Compensator	Compensator			Conditional			Free Rider			Others			Triangular			Unconditional		
	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	−0.016	0.04	0.651	0.128	0.08	0.115	−0.024	0.03	0.369	0.020	0.05	0.701	−0.003	0.03	0.911	−0.105 *	0.06	0.089
Male	−0.005	0.02	0.826	−0.050	0.06	0.407	0.017	0.03	0.545	0.009	0.04	0.843	0.017	0.02	0.446	0.012	0.03	0.674
Married	0.062 **	0.03	0.044	0.353 ***	0.13	0.006	−0.410 ***	0.13	0.002	−0.002	0.06	0.974	0.025	0.02	0.299	−0.028	0.03	0.379
NoReligion	0.011	0.03	0.734	0.051	0.08	0.497	0.029	0.03	0.271	−0.075	0.06	0.234	−0.011	0.02	0.665	−0.006	0.04	0.866
Degree	−0.015	0.03	0.671	−0.023	0.07	0.752	0.005	0.03	0.881	0.046	0.05	0.330	0.030	0.03	0.309	−0.043	0.04	0.285
Age	−0.000	0.00	0.850	0.000	0.00	0.949	−0.005 **	0.00	0.043	0.004	0.00	0.120	0.000	0.00	0.681	0.001	0.00	0.625
Obs									252									
Pseudo R2									0.070									
Prob > F									<0.001									

Note: Multinomial logit regression with robust standard errors. The table reports the average marginal effects of the independent variables for each category of the dependent variable. Obs = observations. Prob > F = p-value of F-test. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10. Unconditional contributions and beliefs.

Sample	Contribution		Belief	
	Mean	Std. Err.	Mean	Std. Err.
Worker (n = 156)	115.82	55.75	101.81	50.45
Student (n = 96)	97.88	60.18	105.71	47.10
Total (n = 252)	108.99	58.02	103.30	49.14

Table A11. Tobit regression (unconditional choice).

	β	Std. Err.	$p > z$
Belief	0.992 ***	0.08	0.000
Student	−22.390 ***	8.52	0.009
Worker	−5.362	7.28	0.462
Male	−10.587	8.79	0.230
Married	−9.910	9.16	0.281
NoReligion	7.958	7.89	0.314
Degree	0.681 *	0.36	0.062
Age	2.074	15.06	0.891
Constant	252		
Obs	−1113.624		
ll	0		
Prob > F	0.992 ***	0.08	0.000

Note: Tobit regression with robust standard errors. Obs = observations. Prob > F = p-value of F-test. * $p < 0.1$, *** $p < 0.01$.

Table A12. Multinomial logit (Strategy method).

Type	Worker	Student	Total
Aggressive	0.64% (n = 1)	2.0% (n = 2)	1.2% (n = 3)
Competitive	7.69% (n = 12)	2.0% (n = 2)	5.6% (n = 14)
Cooperative	4.49% (n = 7)	0% (n = 0)	2.8% (n = 7)
Individualistic	31.41% (n = 49)	55.2% (n = 53)	40.5% (n = 102)
Reciprocating	50.00% (n = 78)	37.5% (n = 36)	45.2% (n = 114)
Others	5.77% (n = 9)	3.1% (n = 3)	4.8% (n = 12)
Total	100% (n = 156)	100% (n = 96)	100% (n = 252)

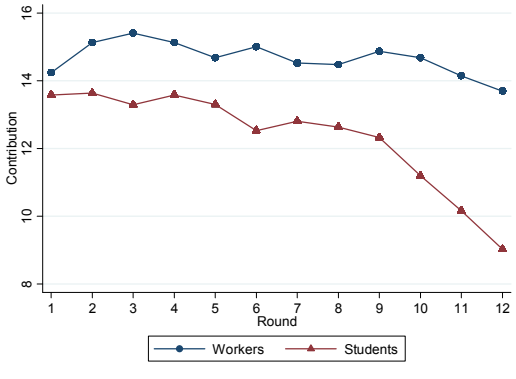


Figure A4. Evolution of average contributions over time.

Table A13. Multinomial logit (Strategy method).

Aggressive	Aggressive			Competitive			Cooperative			Individualistic			Others			Reciprocating		
	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	0.017	0.01	0.134	−0.042	0.06	0.500	−0.369 ***	0.13	0.005	0.231 ***	0.09	0.007	0.026	0.04	0.505	0.137	0.11	0.212
Male	−0.188 *	0.11	0.082	−0.054	0.04	0.193	0.032	0.02	0.104	0.115	0.08	0.157	−0.006	0.03	0.817	0.102	0.08	0.178
Married	−0.185 *	0.11	0.089	0.028	0.04	0.442	−0.021	0.02	0.335	0.047	0.10	0.642	0.007	0.04	0.854	0.124	0.10	0.198
NoReligion	−0.183 *	0.11	0.081	−0.013	0.05	0.813	−0.374 ***	0.13	0.004	0.198 **	0.10	0.044	0.035	0.04	0.323	0.336 ***	0.11	0.002
Degree	0.009	0.01	0.466	−0.001	0.03	0.978	0.001	0.03	0.972	0.091	0.07	0.166	−0.034	0.04	0.386	−0.066	0.07	0.364
Age	0.000	0.00	0.262	−0.001	0.00	0.555	0.002 **	0.00	0.033	-0.004	0.00	0.242	0.001	0.00	0.513	0.002	0.00	0.625
Obs									252									
Pseudo R2									0.085									
Prob > F									<0.001									

Note: Multinomial logit regression with robust standard errors. The table reports the average marginal effects of the independent variables for each category of the dependent variable. Obs = observations. Prob > F = p-value of F-test. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14. Poisson-logit maximum-likelihood hurdle model.

	Hurdle Model 1					
	Decision to Contribute			Decision of How Much to Contribute		
	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	−0.035 **	0.016	0.025	0.020	0.272	0.943
Contribution $t-1$	0.007 ***	0.001	0.000	0.737 ***	0.045	0.000
Contribution $t-2$	0.003 ***	0.001	0.002	0.246 ***	0.027	0.000
Positive deviation $t-1$	−0.015 ***	0.002	0.000	−0.303 ***	0.047	0.000
Negative deviation $t-1$	−0.010 ***	0.002	0.000	0.519 ***	0.112	0.000
Period	−0.009 ***	0.002	0.000	−0.029	0.024	0.223
Male	−0.011	0.013	0.384	0.159	0.211	0.450
Married	0.035	0.027	0.201	−0.115	0.247	0.641
NoReligion	−0.028 *	0.015	0.073	−0.054	0.259	0.834
Degree	0.023 *	0.014	0.095	−0.593 **	0.238	0.013
Age	0.001	0.001	0.473	0.022 *	0.012	0.056
Obs	2520			2292		
	Hurdle Model 2					
	Decision to Contribute			Decision of How Much to Contribute		
	dy/dx	Std. Err.	$p > z$	dy/dx	Std. Err.	$p > z$
Student	−0.025 *	0.015	0.091	−0.023	0.294	0.938
Contribution $t-1$	0.007 ***	0.001	0.000	0.731 ***	0.045	0.000
Contribution $t-2$	0.002 **	0.001	0.014	0.244 ***	0.027	0.000
Positive deviation $t-1$	−0.015 ***	0.002	0.000	−0.304 ***	0.046	0.000
Negative deviation $t-1$	−0.009 ***	0.002	0.000	0.512 ***	0.112	0.000
Period	−0.009 ***	0.002	0.000	−0.031	0.024	0.206
Male	−0.014	0.012	0.219	0.137	0.217	0.528
Married	0.023	0.026	0.362	−0.121	0.248	0.627
NoReligion	−0.017	0.014	0.237	−0.098	0.258	0.704
Degree	0.010	0.014	0.450	−0.548 **	0.238	0.021
Age	0.000	0.001	0.768	0.022 *	0.012	0.079
Beneficent	0.026 *	0.013	0.055	0.006	0.255	0.981
Reciprocating	−0.110 **	0.052	0.036	0.632	0.471	0.180
Individualistic	−0.114 **	0.053	0.030	0.539	0.488	0.269
Competitive	−0.074	0.060	0.218	−0.271	0.690	0.694
Conditional	−0.003	0.017	0.850	0.067	0.263	0.798
Free rider	−0.092 ***	0.025	0.000	−1.077	1.096	0.326
Unconditional	−0.022	0.023	0.340	0.406	0.485	0.403
Obs	2520			2292		

Notes: The decision to contribute is estimated with a logit model with clustered standard errors at individual level. The decision of how much to contribute is estimated with a zero-truncated Poisson model with clustered standard errors at individual level. In the table, we report the average marginal effects of the independent variables. Obs = observations. dy/dx = marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B. Students *vs.* Workers

In this Appendix we present the experimental results obtained by pooling temporary and permanent workers.

B.1. Dictator Game

We can reject the hypothesis that being beneficent is independent of whether a subject is a worker or a student (Chi-squared test, $p < 0.001$). The proportion of beneficent subjects among workers is significantly larger than among students (Chi-squared test, $p < 0.001$).

Table A7 reports the results from the regression analysis. Regression 1 is a Tobit regression where the dependent variable is the amount offered in the Dictator Game; Regression 2 is a Probit regression on the likelihood of being a beneficent subject. In both regressions the explanatory variables include age and dummies identifying the experimental sample (using workers as baseline category), gender (Male = 1 for male subjects), marital status (Married = 1 for married subjects), religious affiliation (NoReligion = 1 for atheist or agnostic subjects), and educational background (Degree = 1 for subjects with a university degree).

B.2. Distribution of Offers in the Dictator Game

Figure A3 shows the distribution of the offers per sample. Comparing the distributions of the offers between workers and students, we reject the null hypothesis that they are the same (Epps-Singleton test, $p < 0.001$).

B.3. Strategy Method

Students and workers significantly differ in the way subjects are classified according to the strategy method (Fisher's exact test, $p = 0.059$). The share of conditional cooperators among students is significantly higher than among workers (Fisher's exact test, $p = 0.043$). Moreover the share of unconditional cooperators is lower among students than workers ($p = 0.010$).

We also ran a multinomial logit regression where the dependent variable is the strategy method classification and where we control for the socio-demographic characteristics of the subjects. For certain categories the sample size is very small, and, therefore, the results should be taken with care. Table A9 reports the marginal effects, for each category of subjects, of this regression. The results confirm that workers are more likely to be unconditional co-operators compared to students.

B.4. Unconditional Contributions and Beliefs in the Strategy Method

If we look at the unconditional contributions in the Strategy Method (Table A10), students contributed on average less than workers. However, the difference is statistically significant only when comparing students with permanent workers (Mann-Whitney test, $p = 0.026$). The beliefs do not significantly differ between workers and students.

Interestingly, the average contributions of workers are statistically higher than their beliefs about the contribution of the others (Wilcoxon signed-rank test, $p < 0.001$), whereas contributions and beliefs of students do not statistically differ (Wilcoxon signed-rank test $p = 0.609$). This result support the evidence that workers are more likely to be unconditional co-operators compared to students, *i.e.*, they contribute no matter what is the contribution of the others.

We also ran a Tobit regression where the dependent variable is the unconditional contribution. Independent variables include dummy variables for students, age, gender (Male = 1 for male subjects), relationship status (Married = 1 for married subjects), religious affiliation (NoReligion = 1 for atheist or agnostic subjects), educational background (Degree = 1 for subjects with a university degree), and beliefs. Table A11 displays the results of the regression. When covariates are controlled for, we observe that Students contribute significantly less than workers. In addition, there is strong evidence that the contribution is positively related to the beliefs about the contribution of the others. This result does not surprise since about 70 % of the subjects were classified as conditional co-operators.

B.5. Decomposed Prisoner's Dilemma

The results of this classification are reported in Table A12. Comparing the proportion of each type of subject between workers and students, the share of Competitive, Cooperative, Reciprocating and Individualistic subjects statistically differ between the two samples (Fisher's exact test, p = 0.049, 0.033, 0.068, and 0.000, respectively).

Two categories collect more than 80% of the subjects in each sample: Individualistic and Reciprocating. Considering these two categories, the shares are unbalanced, as there are more Reciprocating subjects among workers and more Individualistic subjects among students.

We also ran a multinomial logit regression where the dependent variable is the decomposed Prisoner's Dilemma classification and where we control for the socio-demographic characteristics of the subjects. For certain categories the sample size is very small, and, therefore, the results should be taken with care. Table A13 reports the marginal effects, for each category of subjects, of this regression. The results confirm that workers are more likely to be unconditional co-operators compared to students.

B.6. Linear Public Good Game

Figure A4 shows the average contribution in the first 12 rounds of the Linear Public Good Game. Initial contributions do not differ significantly, although we can observe some differences as the game unravels. In the Students sample, average contributions per round and total average contributions are significantly below those

by workers (Mann-Whitney test, $p = 0.085$). Comparing the contributions across the two samples for each round, we reject the hypothesis that the two samples come from the same population only for the last three rounds (Mann-Whitney test, $p = 0.074$, 0.027 and 0.007 respectively). The students' contributions significantly decrease over time (Spearman $\rho = -0.192$, $p < 0.001$), which is a recurrent finding in the literature (see, e.g., [15,16]). Interestingly, there is no significant downward trend for workers (Spearman $\rho = -0.043$, $p = 0.357$), with an average contribution that remains quite constant throughout the 12 rounds of the Public Good Game.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ball, S.B.; Cech, P.A. Subject pool choice and treatment effects in economic laboratory research. *Res. Exp. Econ.* **1996**, *6*, 239–292.
2. List, J.A. Young, selfish and male: Field evidence of social preferences. *Econ. J.* **2004**, *114*, 121–149.
3. Fréchette, G.R. Laboratory experiments: Professionals *versus* students. In *The Methods of Modern Experimental Economics*; Fréchette, G., Schotter, A., Eds.; Oxford University Press: Oxford, UK, 2009.
4. Falk, A.; Meier, S.; Zehnder, C. Do lab experiments misrepresent social preferences? The case of self-selected student samples. *J. Eur. Econ. Assoc.* **2013**, *11*, 839–852.
5. Cadsby, B.C.; Maynes, E. Choosing between a socially efficient and free-riding equilibrium: Nurses *versus* economics and business students. *J. Econ. Behav. Organ.* **1998**, *37*, 183–192.
6. Fehr, E.; List, J.A. The hidden costs and returns of incentives—Trust and trustworthiness among CEOs. *J. Eur. Econ. Assoc.* **2004**, *2*, 743–771.
7. Carpenter, J.; Burk, S.; Verhoogen, E. Workers: The effects of social framing on behavior in Distribution Games. *Res. Exp. Econ.* **2005**, *10*, 261–290.
8. Burks, S.; Carpenter, J.; Goette, L. Performance pay and worker cooperation: Evidence from an artefactual field experiment. *J. Econ. Behav. Organ.* **2009**, *70*, 458–469.
9. Bigoni, M.; Camera, G.; Casari, M. Strategies of cooperation and punishment among students and clerical workers. *J. Econ. Behav. Organ.* **2013**, *94*, 172–182.
10. Harrison, G.W.; List, J.A. Field experiments. *J. Econ. Lit.* **2004**, *42*, 1009–1055.
11. Sala, H.; Silva, J.I.; Toledo, M. Flexibility at the margin and labor market volatility in OECD countries. *Scand. J. Econ.* **2012**, *114*, 991–1017.
12. De Cuyper, N.; de Jong, J.; de Witte, H.; Isaksson, K.; Rigotti, T.; Schalk, R. Literature review of theory and research on the psychological impact of temporary employment: Towards a conceptual model. *Int. J. Manag. Rev.* **2008**, *10*, 25–51.
13. Carpenter, J.; Seki, E. Competitive Work Environments and Social Preferences: Field Experimental Evidence from a Japanese Fishing Community. Available online: http://dx.doi.org/10.2202/1538-0645.1460 (accessed on 12 May 2015).

14. Bellemare, C.; Kroeger, S. On representative social capital. *Eur. Econ. Rev.* **2007**, *51*, 183–202.
15. Andreoni, J. Why free ride?: Strategies and learning in public goods experiments. *J. Public Econ.* **1988**, *37*, 291–304.
16. Croson, R.T.A. Partners and strangers revisited. *Econ. Lett.* **1996**, *53*, 25–32.
17. Brosig, J. Identifying cooperative behavior: Some experimental results in a Prisoner's Dilemma Game. *J. Econ. Behav. Organ.* **2002**, *47*, 275–290.
18. Burlando, R.M.; Guala, F. Heterogeneous agents in public goods experiments. *Exp. Econ.* **2005**, *8*, 35–54.
19. OECD. Taking the measure of temporary employment. In *OECD Employment Outlook*; Organisation for Economic Co-operation and Development: Paris, France, 2002; pp. 127–185.
20. Ichino, A.; Mealli, F.; Nannicini, T. From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *J. Appl. Econom.* **2008**, *23*, 305–327.
21. Booth, A.L.; Francesconi, M.; Frank, J. Temporary jobs: Stepping stones or dead ends? *Econ. J.* **2002**, *112*, F189–F213.
22. Rousseau, D. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*; Sage Publications Inc.: Thousand Oaks, CA, USA, 1995.
23. Martínez, G.; de Cuyper, N.; de Witte, H. Review of temporary employment literature: Perspectives for research and development in Latin America. *Psykhe* **2010**, *19*, 61–73.
24. Engellandt, A.; Riphahn, R.T. Temporary contracts and employee effort. *Labour Econ.* **2005**, *12*, 281–299.
25. Dohmen, T.; Falk, A.; Huffman, D.; Sunde, U. Homo reciprocans: Survey evidence on behavioural outcomes. *Econ. J.* **2009**, *119*, 592–612.
26. Borzaga, C.; Depedri, S. Interpersonal relations and job satisfaction: Some empirical results in social and community care services. In *Economics and Social Interaction: Accounting for Interpersonal Relations*; Gui, B., Sugden, R., Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 125–149.
27. Novkovic, S.; Prokopowicz, P.; Stocki, R. Staying true to co-operative identity: Diagnosing worker co-operatives for adherence to their values. In *Advances in the Economic Analysis of Participatory & Labor-Managed Firms*; Bryson, A., Ed.; Emerald Group Publishing Limited: Bingley, UK, 2012; Volume 13, pp. 23–50.
28. Greiner, B. An Online Recruitment System for Economic Experiments. Available online: http://mpra.ub.uni-muenchen.de/13513/ (accessed on 12 May 2015).
29. Fischbacher, U. Z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
30. Fischbacher, U.; Gächter, S.; Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **2001**, *7*, 397–404.
31. Abeler, J.; Nosenzo, D. Self-selection into laboratory experiments: Pro-social motives *versus* monetary incentives. *Exp. Econ.* **2014**, *18*, 1–20.

32. Pruitt, D.G. Reward structure and cooperation: The decomposed Prisoner's Dilemma Game. *J. Personal. Soc. Psychol.* **1967**, *7*, 21–27.
33. Kuhlman, D.M.; Marshello, A. Individual differences in the game motives of own, relative, and joint gain. *J. Res. Personal.* **1975**, *9*, 240–251.
34. Bigoni, M.; Dragone, D. Effective and efficient experimental instructions. *Econ. Lett.* **2012**, *117*, 460–463.
35. Levitt, S.D.; List, J.A. What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **2007**, *21*, 153–174.
36. Griesinger, D.W.; Livingston, J.W. Toward a model of interpersonal motivation in experimental games. *Behav. Sci.* **1973**, *18*, 173–188.
37. Liebrand, W.B.G. The effect of social motives, communication and group size on behaviour in an *N*-Person Multi-Stage Mixed-Motive Game. *Eur. J. Soc. Psychol.* **1984**, *14*, 239–264.
38. Nikiforakis, N. Punishment and counter-punishment in public good games: Can we really govern ourselves? *J. Public Econ.* **2008**, *92*, 91–112.
39. Bortolotti, S.; Casari, M.; Pancotto, F. Norms of punishment: Experiments with students and the general population. *Econ. Inq.* **2015**, *53*, 1207–1223.
40. Gächter, S.; Herrmann, B. The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *Eur. Econ. Rev.* **2011**, *55*, 193–210.
41. Cardenas, J.C. Groups, commons and regulations: Experiments with villagers and students in Colombia. In *Psychology, Rationality and Economic Behaviour*; Agarwal, B., Vercelli, A., Eds.; Palgrave: New York, NY, USA, 2005; pp. 242–270.
42. Bigoni, M.; Bortolotti, S.; Casari, M.; Gambetta, D.; Pancotto, F. Cooperation hidden frontiers: The behavioral foundations of the Italian North-South divide. In *Quaderni DSE Working Paper No. 882*; Department of Economics, University of Bologna: Bologna, Italy, 2013.
43. Hoffman, E.; McCabe, K.; Smith, V.L. Social distance and other-regarding behavior in Dictator Games. *Am. Econ. Rev.* **1996**, *86*, 653–660.
44. Leary, M.R.; Kowalski, R.M. Impression management: A literature review and Two-Component Model. *Psychol. Bull.* **1990**, *107*, 34–47.

The Loser's Bliss in Auctions with Price Externality

Ernan Haruvy and Peter T. L. Popkowski Leszczyc

Abstract: We consider auctions with price externality where all bidders derive utility from the winning price, such as charity auctions. In addition to the benefit to the winning bidder, all bidders obtain a benefit that is increasing in the winning price. Theory makes two predictions in such settings: First, individual bids will be increasing in the multiplier on the winning price. Second, individual bids will not depend on the number of other bidders. Empirically, we find no evidence that increasing the multiplier increases individual bids in a systematic way, but we find that increasing the number of bidders does. An analysis of individual bidding functions reveals that bidders underweight the incentives to win and overweight the incentives to lose.

Reprinted from *Games*. Cite as: Haruvy, E.; Popkowski Leszczyc, P.T.L. The Loser's Bliss in Auctions with Price Externality. *Games* **2015**, *6*, 191–213.

1. Introduction

Auctions play a prominent role in charity fundraising, with billions collected through hundreds of thousands of silent and live auctions in the US alone [1]. Internet auctions with price externality in particular are on the rise, with eBay dedicating a special section for charity auctions. Not only do these auctions play an important role for non-profit organizations, but an increasing number of for-profit firms are sponsoring charity events to establish themselves as good corporate citizens. In turn, sponsorship of social causes may improve a firm's image or profits [2–4][1].

Individuals bidding in charity auctions are assumed to care about the cause for which they are bidding. Thus, in the literature on auctions with price externality, charitable bidders are typically modeled as maximizing an objective function in which they receive additional utility that is increasing in the final price paid in the auction (*i.e.*, the amount of money going to charity). This is equivalent to an auction where the losing bidder cares about the price paid by the winning bidder [11]. We refer to this class of auctions as *auctions with price externality*.

[1] There is a growing body of research in marketing that has concluded that linking product purchases with donations to charities has a positive impact on perceptions [4–7]. However, these positive effects are not universal, since several researchers have shown that in certain instances it may lead to a reduction in purchase intention [8–10].

In the present work, we specify laboratory payoffs corresponding to the functional form of auctions with price externality [12]. That is, bidders in auctions with price externality receive extra utility from the selling price of the auction, expressed as a proportion of that price. We refer to this proportion as the *multiplier* (henceforth also referred to as alpha). We study settings where both winning and losing bidders receive such a proportion of the winning price, and this multiplier varies across conditions.

Theoretical predictions in auctions with price externality [13,14] can be stated in the form of two hypotheses: (1) Revenues will increase in the multiplier, and (2) optimal bids will not change in the number of other bidders. We see that despite having an experimental setup that closely corresponds to the theoretical setting bidders do not in general conform to these predictions. We specifically show that decisions were affected by theoretically irrelevant considerations [15].

First, we document a persistent underbidding pattern in auctions with few bidders. In three-bidder auctions, we see that bid levels decline in response to changes in the multiplier—In a direction opposite to the direction prescribed by theory[2].

Second, we find that in contrast to theoretical predictions, the number of bidders affects the extent of underbidding for a higher multiplier. A larger number of bidders may mitigate and even reverse the above pattern.

Third, in English auctions, the tendency to underbid in response to higher multipliers is not merely driven by a misalignment of bids with valuations. Rather, higher multipliers appear to result in greater reluctance to react to competitors' bids. We therefore highlight a particular source of inconsistency between the theory and the data—an apparent tendency by bidders to bid as if they underweight the incentive to win and overweight the incentive to lose in the charity setting.

2. Theory

2.1. Setting and Theoretical Properties

An auction with price externality is defined as an auction in which bidders receive a utility that increases in the winning price for the auction. We study two auction formats. The first format is the second price sealed bid (hereafter SPSB). In

---

[2] Isaac *et al*. (2010) [12] encountered a hint of this effect in second price sealed bid auctions. They found that charitable preferences of the type investigated here do not significantly increase individual bids or auction revenues, thus identifying a puzzle. In their carefully designed second price sealed bid experiment, four decision makers played 40 auction rounds with multipliers of 0 in a benchmark condition, 0.15 in a low bonus condition or 0.50 in a high bonus condition (different decision makers in each condition). From their plots, bids in the 0.5 multiplier (second price basic charity) condition are on average lower than theoretical predictions, and this gap is quite substantial.

SPSB, each bidder submits a single bid simultaneously with all other bidders. The bidder who submits the highest bid wins the auction and pays the second highest bid.

The second format we investigate is the English auction, wherein bidders may place ascending bids throughout the auction. It is optimal for bidders to bid up to their valuation in English auctions, and revenues will be identical to a second price sealed-bid auction in the absence of major jump bids [13,14,16].

For both SPSB and English auctions with price externality, the winning bidder receives his valuation for the item and pays the winning price. In addition, each bidder, whether he won or lost, receives a constant portion α of the winning price [12–14].

We derive (see Appendix A) the optimal bid in this environment. Three theoretical properties emerge from this derivation.

Property 1. *The bids in auctions with price externality are increasing in the multiplier (α).*

Property 2. *Auctions with price externality are efficient.*

Property 3. *The bids in auctions with price externality do not depend on the number of bidders (n).*

The first property is straightforward as a higher multiplier results in greater utility, and therefore should result in higher bids. Efficiency, indicated in Property 2, is the property that the bidder with the highest valuation for the item will win the item. The third theoretical property seems to contradict findings from empirical research. Empirical findings often indicate that competitive intensity increases with more bidders [17–19].

In our study, *bidder valuations* are induced and therefore exogenous. *Willingness to pay* (*WTP*) is the maximum amount a bidder is willing to pay in an auction and this is the value we measure in our study. However, WTP will differ from these bidder valuations in auctions with price externality, because there are added incentives to win and to lose, as detailed in Appendix A.

Bids in the English auction format cannot be looked at in the same way as bids in the SPSB format. In the SPSB format, all bids are equally informative about bidder WTP, whether or not the bidder ends up winning or losing the auction. Under SPSB, it is a strictly dominant strategy for each bidder to bid his true WTP. Thus, we can view the observed bid as a proxy for WTP. In contrast, in an English auction, each bidder should bid incrementally as long as his WTP has not been reached. As such, if all bidders bid incrementally, the highest bid of each losing bidder is an indication of WTP. If there are *jump bids*—Bids that are more than an increment above the previous bid—then the last bid by each bidder is a less informative statistic. Therefore, this is at best a noisy indication of WTP. In the English auction analysis, we therefore focus

on losers' bids. In particular, we pay attention to a construct we call *loser underbidding*, where bidders bid below their theoretically predicted bid.

2.2. Hypotheses

We now state three important hypotheses arising directly from the established theory. The findings, to be reported shortly, will pertain directly to the hypotheses. As we alluded to in earlier sections, the findings of this work offer evidence to suggest rejection of most of these hypotheses.

2.2.1. Hypothesis 1. Individual Bids and Revenues Increase in the Multiplier

As the extra utility that bidders receive from the selling price of the auction, increases we expect an increase in bids and revenue. The intuition behind this hypothesis is that the multiplier serves as a subsidy to the winning bidder, who gets a portion (equal to the multiplier) of his winning price refunded. As the multiplier increases, the subsidy to the winning price is higher, resulting in an incentive to bid higher. As bids increase, so does the revenue.

It is a generalization that is in line with Equation (4) in Appendix A and is consistent with theoretical results elsewhere [13,14]. Empirically, there is conflicting evidence for this hypothesis. On the one hand [12], which has a design similar to the present one, do not find support for this prediction in SPSB. On the other hand, field experiments by [20,21] report that bidders in English auctions are willing to pay a higher price for higher donation promises. Both papers ran simultaneous pairs of English auctions identical in all respects but the percentage of proceeds donated to charity. While these papers did not directly vary the multiplier, higher donations to charity should result in greater utility to charitable bidders.

We further note that the optimal bid does not vary in the number of bidders, which is a well-known theoretical result, consistent with Equation (4) in Appendix A and with results elsewhere in the literature [14]. We express this as hypothesis 2.

2.2.2. Hypothesis 2. Auctions with Price Externality are Efficient

Consistent with the theory for auctions without price externalities, it is expected that with price externalities the bidder with the highest valuation will win the auction. As winning bidders receive a subsidy proportional to the multiplier, they have an incentive to bid higher with a price externality. Therefore, the bidder with the highest valuation has the greatest incentive to win the auction.

We further note that the optimal bid does not vary in the number of bidders, which is a well-known theoretical result, consistent with Equation (4) in Appendix A and with results elsewhere in the literature [14]. We express this as hypothesis 3.

2.2.3. Hypothesis 3. Bidders' WTP is Unchanging in the Number of Other Bidders

The theoretical prediction of unchanging WTP in response to the number of bidders notwithstanding, empirical studies typically find that the number of bidders in an auction does impact bidder behavior. For example, [19] report that auctions with a higher number of bids resulted in a higher WTP. In contrast, [22] find that bidders in auctions with fewer bidders tend to bid more aggressively. Unlike the competitive settings in these other works, in the present setting of auctions with price externality each bidder's utility increases with higher selling price, even when that bidder is not the one paying the price. The number of other bidders directly affects the probability of not paying the price. Specifically, as the number of bidders increases, the likelihood of winning decreases. Having more opponents enables a bidder to increase his expected utility by raising the eventual price paid by others, while maintaining a low probability of winning. In summary, when there are more bidders we expect to see more aggressive bidding. This is related to the concept of shill bidding, which is bidding with the intent of raising the price without winning, and we may expect to see more shill bidding in auctions with price externality [21].

3. Experimental Design

In total, 570 undergraduate students at a major North American public university participated in the study. They were recruited for an auction experiment but were not told the purpose of the study. Participants were invited in groups of 12 to 18, and groups were randomly assigned to one of 12 different conditions in a 2 (auctioned format: English *vs.* SPSB) $\times$ 2 (number of bidders participating in the auction: 3 *vs.* 6) $\times$ 3 (multiplier: 0 *vs.* ¼ *vs.* ¾) between-subjects design. The number of bidders was either three bidders per group or six bidders per group, depending on the condition. The number of participants by condition is summarized in Table 1. Each session tested a single experimental condition, and bidders participated in ten rounds of auctions per session.

Table 1. Different experimental conditions and number of sessions in each condition.

		Multiplier		
Format	**# Bidders**	**0**	**¼**	**¾**
SPSB	**3 bidders**	*"3-bidder m = 0"* 15 sessions	*"3-bidder m = ¼"* 19 sessions	*"3-bidder m = ¾"* 13 sessions
SPSB	**6 bidders**	*"6-bidder m = 0"* 4 sessions	*"6-bidder m = ¼"* 6 sessions	*"6-bidder m = ¾"* 5 sessions
English	**3 bidders**	*"3-bidder m = 0"* 13 sessions	*"3-bidder m = ¼"* 25 sessions	*"3-bidder m = ¾"* 17 sessions
English	**6 bidders**	*"6-bidder m = 0"* 6 sessions	*"6-bidder m = ¼"* 15 sessions	*"6-bidder m = ¾"* Sessions

Participants were seated next to private computer terminals and were given written instructions (included in Appendix B). After decision makers were given an opportunity to review the written instructions, the instructions were read aloud by the experimenter. Decision makers were then allowed to ask questions. This was followed by two practice auctions and another round of questions. Decision makers were not allowed to talk during the study and had to remain seated at the computer terminal until the completion of the study. The experiment was computerized, using the popular experimental platform of zTree [23].

The private valuation for the item, denoted by v was uniformly distributed between 50 and 100, and restricted to integers. New values were drawn for each bidder in each of the rounds decision makers participated in. The item was denoted generically as "Item A," and bidders were shown their private value for the item in large letters on the screen. Hence, bidders knew their own valuation, but the only information they had about other bidders' valuations was that they were randomly drawn from 50 to 100. Bidders received pay in the form of tokens. The exchange rate ranged from 1 to 15 cents per token[3].

The multiplier, denoted by α in Equation (1), took on the value of 0, ¼ or ¾. When the multiplier was 0, each winning bidder received his private valuation minus the winning bid for each round in which he won and the losing bidders received nothing. With a positive multiplier, all bidders, whether they won or lost, received in each round an additional amount equal to the winning bid times the multiplier. Thus, losing bidders received a portion of the winning bid whereas winning bidders can be thought of as receiving a subsidy so that they pay only a fraction of their winning bid. This approach is similar to the method used by [12]. They showed that results of this approach were consistent with those of an actual auction with price externality where proceeds were donated to charity.

In conditions that involved the English auction format, each round lasted for a minimum of two minutes, with a "soft" closing time—A 15 s extension occurred when a last second bid was placed. Thus, for example, if there are 3 s remaining and a bid is placed, the timer will reset and there will be 15 s remaining from that point on. The purpose of that feature was to avoid sniping—Or last second bids.

[3] The intent was to calibrate earnings to be roughly the same for different sessions—in the $15 range per subject including a show up fee. We varied the exchange rate within treatments to test for exchange rate effect. No exchange rate effect was found.

4. Experimental Results

Table 2 provides a comprehensive overview of our results in terms of efficiency, revenue and predicted bids, while Table 3 presents the significance tests of these results, comparing outcomes for different conditions (multipliers and number of bidders). Table 4 provides a summary with the findings from our hypotheses tests.

4.1. The Effect of the Multiplier on Individual Bids and Revenues

We first consider hypothesis 1 which predicts that individual bids and revenues increase in the multiplier. We look at the predictions for individual bids, and compare bids to theoretical predictions as well as the impact of the multiplier on auction revenue.

We begin with the analysis of the SPSB format, because in that format, we observe the uncensored bids by all bidders, which is essential for the first part of hypothesis 1 pertaining to comparisons of individual bids.

The left-hand side of Figure 1 shows the bids prescribed by the theory for each possible bidder valuation under the SPSB conditions. The theory-predicted bids are linear in valuation for each of the conditions. The theory unambiguously prescribes that the bid-valuation relationship, for all but the top of the support, would shift upwards as the multiplier increases. That means that for a given valuation, a higher multiplier would imply a higher bid for most of the range of valuations. The lines intersect at the top of the valuation support. The right-hand side of Figure 1 shows the estimated linear bid functions that are based on the actual bids, which seem to indicate less of an impact due the multiplier than suggested by theory.

WTPs obtained from the auctions for 3-bidder SPSB are different from the theoretical predictions. The p-values are 0.01, 0.33, <0.01, for the 3-bidder $m = 0$ (control) condition, $m = ¼$ condition and $m = ¾$ condition, respectively. Results for 6-bidder SPSB do not differ from theoretical predictions (p-values are 0.62, 0.46, 0.10, for 6-bidders $m = 0$ condition, $m = ¼$ condition, and $m = ¾$ condition respectively). Unlike the clearly separated predictions for the different multipliers, the estimated linear bid functions are remarkably close to one another for the range of values used in the study. The intercepts for the ¼ and ¾ multiplier are not significantly different from the intercept for the $m = 0$ condition ($p = 0.80$ for condition ¼ and $p = 0.12$ for condition ¾). The slopes for the ¼ and ¾ multiplier are not significantly different from the slope for $m = 0$ ($p = 0.46$ for condition ¼ and $p = 0.35$ for condition ¾). Thus, the data ranking of the condition manipulations is inconsistent with the theoretical prediction for the three-bidder conditions, although it appears in line with theoretical predictions for six player conditions.

Table 2. Comparison of SPSB and English auctions for different conditions (multipliers and number of bidders) (Standard errors in parentheses).

	Results SPSB Auctions				Results English Auctions			
Condition	**1. Predicted Bid Minus Bid [a]**	**2. Valuation Minus Bid [a]**	**3. Revenue**	**4. Efficiency [b]**	**5. Predicted Bid Minus Bid [a]**	**6. Valuation Minus Bid [a]**	**7. Revenue**	**8. Efficiency [b]**
3-bidder, $m = 0$	$−2.90 (1.43)	$−2.90 (1.43)	$93.44 (1.35)	63.80 (5.10)	$13.24 (2.11)	$13.24 (2.11)	$75.13 (0.82)	84.28 (4.68)
3-bidder, $m = ¼$	8.45 (3.32)	3.49 (3.32)	95.77 (3.68)	56.93 (4.51)	22.60 (1.99)	16.58 (2.02)	71.49 (1.33)	64.57 (3.73)
3-bidder, $m = ¾$	14.66 (3.35)	3.65 (3.29)	90.61 (3.38)	65.95 (4.99)	27.75 (2.17)	14.66 (2.16)	78.33 (0.71)	69.50 (4.06)
6-bidder, $m = 0$	0.70 (5.30)	0.70 (5.30)	101.65 (3.79)	55.00 (11.90)	23.92 (1.36)	23.92 (1.36)	86.17 (1.52)	68.33 (7.17)
6-bidder, $m = ¼$	5.53 (4.79)	0.572 (4.83)	110.37 (7.54)	36.67 (6.15)	24.48 (1.29)	19.34 (1.28)	96.35 (0.88)	49.33 (4.09)
6-bidder, $m = ¾$	−8.01 (5.09)	−18.56 (5.03)	151.50 (14.96)	30.00 (4.47)	8.83 (8.90)	−2.72 (8.97)	128.44 (20.03)	38.75 (8.95)

[a] In English auctions, this statistic pertains only to losers' bids; [b] % of time highest value bidder won.

Table 3. T-tests for significance of different outcome from Table 2, for different conditions (multipliers and number of bidders). (T-statistics and *p*-values reported).

	Results SPSB Auctions				Results English Auctions			
Test	**1. Predicted Minus Bid [a]**	**2. Valuation Minus Bid [a]**	**3. Revenue**	**4. Efficiency [b]**	**1. Predicted Bid Minus Bid [a]**	**2. Valuation Minus Bid [a]**	**3. Revenue**	**4. Efficiency [b]**
T-test for: 3 bidders 0 *vs.* ¼	T = 2.87 *p* = 0.005	T = 1.62 *p* = 0.109	T = 0.54 *p* = 0.590	T = 1.01 *p* = 0.315	T = 2.54 *p* = 0.010	T = 0.79 *p* = 0.180	T = 1.70 *p* = 0.090	T = 3.18 *p* = 0.002
T-test for: 3 bidders¼ *vs.* ¾	T = 1.27 *p* = 0.207	T = 0.03 *p* = 0.974	T = 0.98 *p* = 0.328	T = 1.32 *p* = 0.190	T = 1.90 *p* = 0.060	T = 0.46 *p* = 0.650	T = 4.08 *p* < 0.001	T = 1.12 *p* = 0.270
T-test for: 3 bidders0 *vs.* ¾	T = 5.06 *p* < 0.001	T = 1.91 *p* = 0.059	T = 0.82 *p* = 0.416	T = 0.30 *p* = 0.766	T = 4.32 *p* < 0.001	T = 0.42 *p* = 0.673	T = 2.80 *p* = 0.007	T = 2.28 *p* = 0.025
T-test for: 6 bidders 0 *vs.* ¼	T = −0.66 *p* = 0.527	T = 0.02 *p* = 0.987	T = 0.88 *p* = 0.404	T = 1.51 *p* = .170	T = 0.11 *p* = 0.913	T = 0.90 *p* = 0.377	T = 3.74 *p* = 0.001	T = 3.14 *p* = 0.005
T-test for: 6 bidders¼ *vs.* ¾	T = 1.93 *p* = 0.086	T = 2.73 *p* = 0.023	T = −2.59 *p* = 0.029	T = 0.84 *p* = 0.421	T = 2.05 *p* = 0.053	T = 2.88 *p* = 0.009	T = 2.21 *p* = 0.038	T = 1.31 *p* = 0.205
T-test for: 6 bidders 0 *vs.* ¾	T = 1.17 *p* = 0.279	T = 2.61 *p* = 0.035	T = −2.88 *p* = 0.024	T = 2.15 *p* = 0.068	T = 1.43 *p* = 0.179	T = 2.51 *p* = 0.028	T = 1.81 *p* = 0.096	T = 2.75 *p* = 0.018

[a] In English auctions, this statistic pertains only to losers' bids; [b] % of time highest value bidder won.

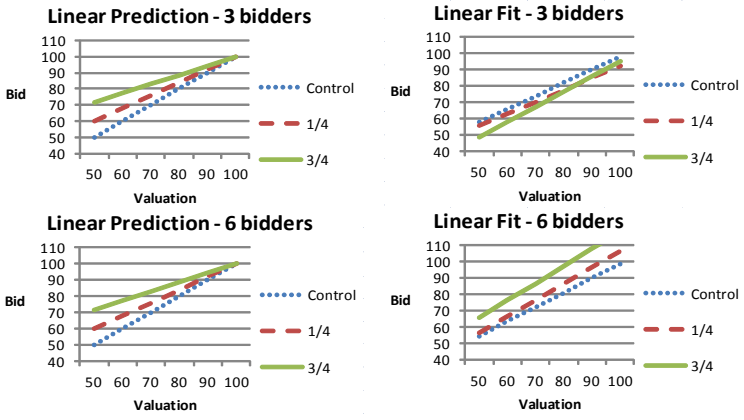


Figure 1. SPSB conditions. Theory and linear fit of the bid data on valuation.

Next, we examine how (under) bidding patterns correspond to bidder valuations in the English auction format. Unlike the SPSB format, in the English auction format we can only gauge underbidding for losing bidders. We find that underbidding in English auctions with $n = 3$ significantly increases ($p = 0.01$) between 0 and ¼ and marginally increases ($p = 0.06$) between ¼ and ¾. For $n = 6$ English auctions, we see no evidence of increased underbidding with higher multiplier[4]. In contrast, for $n = 6$, the magnitude of underbidding decreases in the multiplier between $m = ¼$ and $m = ¾$, and this is significant ($p = 0.05$).

One important finding is that there is substantial loser underbidding (even in the control condition with zero multiplier)[5]. Such underbidding has been documented before [24] and explained with the observation that bidders are reactive and inexperienced [24–26].

Figure 2 shows the extent of underbidding over time—across the auctions in which decision makers participated—separately for three bidders and six bidders. Especially in the SPSB formats—less so in the English formats—we observe a decrease in the extent of underbidding over time. This greater movement overtime in SPSB auctions than in English auctions is likely due the fact that in English auctions bidders are able to adapt to what the competition does in the course of the auctions

---

4 Note that revenue (equal to the winning price) in condition $m = ¾$ with 6 bidders is significantly above the maximum bidder valuation of 100, which is nevertheless below the theoretical prediction for this condition.

5 Some underbidding in the control treatment is expected. The predicted bid for a losing bidder is at one's valuation. Any bid above one's valuation in the zero-multiplier condition results in a loss conditional on winning regardless of bidding strategy or beliefs. So any bidder deviation from prediction should be below the predicted bid. Deviations could be due to any number of reasons including bidders bidding intermittently, jump bidding, or simply truncated errors.

(within-round learning), whereas in SPSB auctions learning can only take place between rounds.

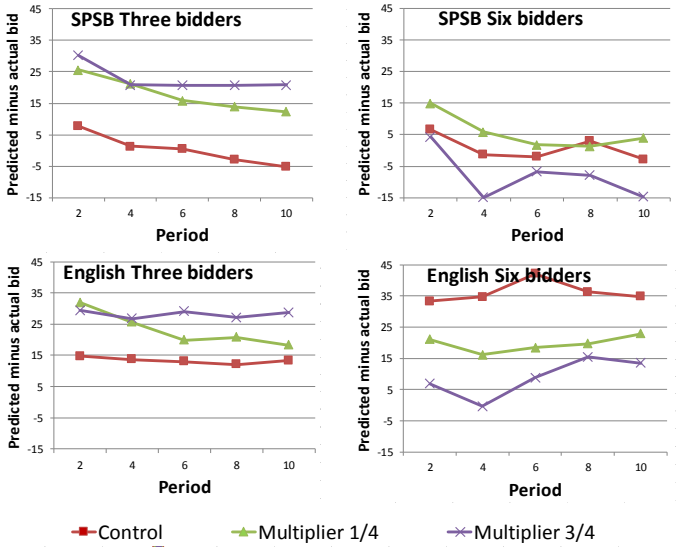


Figure 2. Extent of Underbidding over time in SPSB and English Auctions.

In English auctions with three bidders, decreased underbidding over time is only evident for the 3-bidder m = ¼ condition. For the 3-bidder m = ¾ condition, losing bidders gain more and there is less incentive to adapt bidding behavior and so we see flatter curves in the m = ¾ condition. This persistence of underbidding is reminiscent of results in the auction and winner's curse literature regarding the persistence of overbidding over time in common value auctions [27] as well as takeover bids [28][6].

In 6-bidder English auctions with m = 0 we see a small increase and then decrease in underbidding over time. For both m = ¼ and m = ¾ conditions we initially see a decline in underbidding followed by an increase. In 6-bidder SPSB auctions, underbidding is less pronounced in the m = 0 and m = ¼ conditions. In the 6-bidder m = ¾ condition, we first see movement towards less underbidding, and eventually towards overbidding.

We construct a linear regression of underbidding on bidder valuation, such that the dependent variable is underbidding (either loser valuation minus bid; or

[6] Foreman and Murnighan (1996) [27] investigated whether feedback and relevant experience over four weeks could help alleviate overbidding in common value auctions (known as the winner's curse). They found that while some learning occurred, the overbidding was remarkably persistent over time.

theoretically predicted best bid minus bid) and the explanatory variable is bidder valuation, plus a constant. The resulting fitted lines are exhibited in Figure 3.

The figures point to the same relative rankings. In the 3-bidder English auction conditions, the 3-bidder m = ¾ English auction condition exhibits the greatest underbidding, followed by m = ¼ English auction condition and then the 3-bidder m = 0 English auction condition. However, in the m = 0 condition and the 3-bidder m = ¾ condition bidders exhibit a negative slope of underbidding to valuation as shown in Figure 3: Underbidding drops as bidders' valuations increase. This is intuitive—as bidders have a higher chance of winning, their incentives to compete increase.

The 3-bidder m = ¾ condition's high intercept in Figure 3 could mean that likely losers (low valuation bidders) in the ¾ condition are more discouraged relative to the other conditions. Note however, that we do not observe this pattern for the 6-bidder conditions where we expect that bidders should be even more discouraged, due to reduced likelihood of winning.

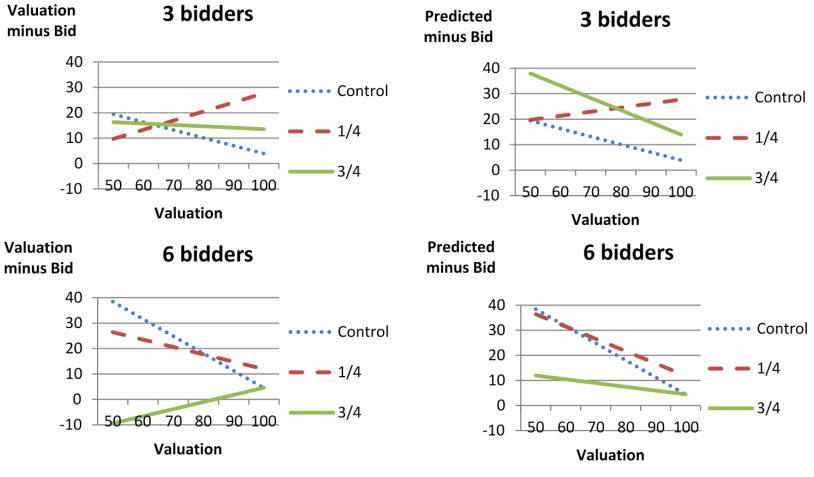


Figure 3. English auctions. Relationship between loser underbidding and valuation.

The 3-bidder m = ¼ condition has a highly significant ($p < 0.001$) positive slope of underbidding in valuation. That is, as valuations increase, bidders are more likely to underbid. This suggests some reluctance to win in that condition because as one's likelihood to win increases, his underbidding is more pronounced.

In the 6-bidder English auction conditions, the m = ¼ condition was shown to have more underbidding than the 6-bidder m = ¾ condition (see bottom panel of Figure 3). Figure 3 confirms that ranking. Recall that we speculated early on that there might be less underbidding in the 6-bidder conditions because the probability

of winning is lower. The 6-bidder $m = ¾$ condition shows a small positive slope in the theoretical underbidding (but a small negative slope in the underbidding relative to valuation). Neither slope in the 6-bidder $m = ¾$ condition is significant ($p > 0.10$).

In summary, we observe that underbidding increases as the multiplier increases, which is a rejection of hypothesis 1. In particular, English auctions with three bidders display a clear rejection of hypothesis 1 in that revenues fall and the magnitude of underbidding increases in the multiplier. These patterns parallel the patterns in SPSB, offering a consistent set of evidence.

4.2. The Effect of the Multiplier on Auction Revenues

The second part of hypothesis 1, pertaining to revenue comparisons, is assessed in column 4 of Table 2. Revenues in the three-bidder SPSB are 93.44 (1.35), 95.77 (3.68), and 90.61 (3.38) for condition $m = 0$, condition ¼ and condition ¾ respectively. The t-tests reported in Table 3 show no significant differences in revenues between the different multipliers in the three-bidder SPSB conditions. Thus, our tests indicate that the multiplier does not significantly affect revenues in the three-bidder SPSB conditions. As we will see in subsequent analysis, this is due to the disproportional impact of the multiplier on bidders' reactions to the incentive to lose.

In the six-bidder SPSB conditions, the revenues are clearly increasing and are 101.65 (1.37), 110.37 (3.79) and 151.50 (14.96) for condition $m = 0$, condition ¼ and condition ¾ respectively. Thus, Table 3 indicates that the multiplier significantly affects revenues in the six-bidder SPSB conditions and in the direction predicted by the theory. Table 3 provides pairwise tests between the conditions.

For English auctions revenue comparisons show that for $n = 3$, a higher multiplier does not monotonically result in higher prices as theory predicts. In comparing the $m = 0$ condition to the ¼ multiplier condition for $n = 3$, we see, in contrast to hypothesis 1, a marginally significantly *lower* revenue with the multiplier increase ($p = 0.09$) (see Table 3 for significance tests). However, for $n = 6$, revenues increase as the multiplier increases from 0 to ¼, in accordance with the theory ($p = 0.001$). Moreover, we find that as the multiplier increases from ¼ to ¾, revenues significantly increase for both $n = 3$ ($p < 0.001$) and for $n = 6$ ($p = 0.038$), in support of hypothesis 1. We also see for $n = 3$ a significant ($p = 0.007$) difference in revenues between the 0 multiplier and the ¾ multiplier in the direction predicted by hypothesis 1. This difference is only marginally significant for $n = 6$ ($p = 0.096$).

The main finding with respect to hypothesis 1, that revenue increases in the multiplier, is that we find support for this in the English auction, but only partial support in the SPSB auction (only for the six-bidder condition). Thus, the theory regarding the multiplier does not seem to hold all the way for smaller groups of bidders.

4.3. Efficiency in Auctions with Price Externalities

Hypothesis 2 proposed that auctions with price externality are efficient. Recall from property 2 that efficiency implies that the highest valuation bidder wins the auction. In Table 2, efficiency is computed as the percentage of time that the highest valuation bidder won. Ideally, according to Property 2, efficiency should be at 100%. Table 2 shows that the loss of efficiency (see columns 4 and 8) resulting from underbidding is pretty substantial for both SPSB and English auctions. In addition, as expected, efficiency decreases as the number of competing bidders increases from three to six, and efficiency is greater for English than for SPSB auctions.

In general, efficiency is higher in auctions without externalities (a multiplier) than for auctions with a multiplier. For English auctions this is the case for both positive multipliers and for both 3 and 6 bidder-auctions (see Table 3), while for SPSB auctions efficiency is only significantly lower, relative to $m = 0$, for the ¾ condition with 6 bidders.

Overall, auctions with externalities are not efficient as in less than half of the instances the highest bidder actually wins in a SPSB auction and only a little above half of the instances in English auctions. While in general English auctions were more efficient, as bidders can update their bids, efficiency is more influenced by the multiplier (a reduction in efficiency) in the English auctions than in the SPSB auction.

4.4. The Effect of the Number of Bidders on WTP

Hypothesis 3, predicted that bidders' WTP is unchanging in the number of other bidders. For SPSB auctions we find that increasing the group size from three to six bidders yields a significant increase in individual bids, which is counter to what theory prescribes (see Column 1 of Table 2). Thus, we reject hypothesis 3 for the SPSB format and conclude that increasing the number of bidders appears to result in higher individual bids.

In the English auction we find significant underbidding for losing bidders (see Column 1 of Table 2). We see that the losing bidders typically underbid relative to their theoretically prescribed bids but that the deviation is significantly smaller for 6 bidders with a multiplier of ¾ relative to 3 bidders with a multiplier of ¾. This goes against the theoretical result as expressed in hypothesis 3. We can thus reject hypothesis 3 for the ¾ multiplier, but not for the ¼ multiplier.

Overall we found limited support for the hypotheses, derived from theory. In particular, we observe significant underbidding in both English and SPSB auctions, which is contrary to theoretical predictions. We also find, contrary to theoretical predictions, that auctions with price externality (both English and SPSB) are not efficient. Finally, in contrast to theoretical predictions, we find that bidders' WTP tends to vary with the number of bidders. The results of these hypotheses are summarized in Table 4.

Table 4. Summary of Hypotheses tests for different conditions.

	H1. Individual Bids Increase in the Multiplier	**H1. Revenues Increase in the Multiplier.**	**H2. Auctions with Price Externality are Efficient.**	**H3. Bidders' WTP is Unchanging in Number of Other Bidders.**
SPSB–3 bidders	Not Supported	Not Supported	Not Supported (same as $m = 0$)	Not Supported
SPSB–6 bidders	Supported (only for $m = ¾$)	Supported (only for $m = ¾$)	Not Supported (same as $m = 0$)	
English–3 bidders	Not Supported	Supported (only for $m = ¾$)	Not Supported (less efficient then $m = 0$)	Supported (for ¼ multiplier) Not Supported (for ¾ multiplier)
English–6 bidders	Supported (only for $m = ¾$)	Supported	Not Supported (less efficient then $m = 0$)	

4.5. Analysis of Heterogeneous Population Models

Up to this point we have assumed that bidders are homogeneous in their bidding behavior. However, it is to be expected that bidders vary in their behavior. In particular, we expect that some bidders will be less aggressive and others overaggressive in their winning propensity (independent of the experimental conditions). Note that if no bidder is overaggressive or under-aggressive, we should expect to see in the three bidder auctions, 3.33 wins per bidder out of the ten auctions that bidders participate in. This is in sharp contrast to the actual proportions observed in the data, which indicate a significant amount of both overbidding and underbidding. We do not observe a specific trend in the data, suggesting that this is due to individual differences rather than due to the different conditions in our study (*i.e.*, the multiplier or the number of bidders).

The existence of both overbidding and underbidding, suggests that we ought to be using a utility-based econometric estimation that permits bidder heterogeneity. The utility function we employ is linear in its terms (for a more formal presentation, see Appendix A). The first term in the utility is the direct payoff—the difference between the private value for the item and the bid—multiplied by the probability of winning with that bid. The next two terms are additional payoffs winning and losing bidders receive from the money raised. Both winners and losers receive an additional payoff α_i from each dollar of revenue. Following [12,13], we allow each bidder to possess two parameters α and β. [12] describe the parameters α and β as denoting the utility a bidder gains for each additional unit of revenue the auctioneer raises when the bidder loses the auction and when the bidder wins the auction, respectively[7]. We refer to parameter α as the *incentive to losers* and to parameter β as the *incentive to winners.*

[7] Bidders who gain greater utility from seller revenue when they are the winning bidder could be interpreted as wanting to be seen as a generous person. Salmon and Isaac (2006) [13] and Isaac *et al.* (2010) [12] refer informally to such preferences as preferences to see-and-be-seen.

We use the mixture model approach for estimation. The mixture model approach posits discrete segments of bidders, where bidders within a segment all possess the same underlying behavioral parameters. The point of this mixture model is to group bidders into "types". This type of model is often very useful in making predictions or characterization in a setting such as auctions with price externality.

In Table 5 we report estimates of the three-segment bidding functions for the SPSB and English conditions, using the structural functional form of Equation (4′). For the purpose of estimation, all bidders are used for the SPSB bidding functions (winners and losers) as opposed to only losing bidders for the English auction estimates. This is because in SPSB, all bidders have symmetric information and symmetric strategies.

Several things are worth noting in Table 5. First, we see that the number of segments differs across conditions. The number of segments to include is based on likelihood ratio tests, incrementally increasing the number of segments till there is no further improvement in fit. For identification, when there are three segments, one of the three segments' alpha and beta parameters are fixed at zero. This parameterization cannot be rejected in terms of likelihood and fixing the parameters gets us cleaner inference on the remaining parameters.

We see that for segment 1 in the one-fourth condition, the coefficients for alpha and beta are not individually significantly different from 0. However, these two parameters are jointly significantly different from zero at $p < 0.05$, and the difference between them is different from zero at $p < 0.05$.

Other than the three-segment characterization, a key take-away is that in the $m = 0$ condition as well as the one-fourth condition, there are both a losing-prone segment (segment 1) and a winning-prone segment (segment 2). This is especially notable for the $m = 0$ condition, because in that condition, the multiplier is zero so there should not be an externality to either losers or winners so for rational bidders, alpha and beta should be 0.

The alpha and beta parameters in all conditions are significantly different from one another, suggesting that the incentive to losers (alpha) and incentive to winners (beta) are not perceived as being of the same magnitude, which is in contrast to the actual underlying incentives. It is interesting that the underlying parameters are generally decoupled from the incentive structure. That is, the monetary incentives influence the underlying propensity to win or lose but we cannot detect any clear and meaningful pattern of that relationship. This seems to suggest that bidder aggressiveness (both over-aggressiveness and under-aggressiveness) is bidder-specific and not necessarily related to the incentive structure. The three-fourth SPSB three-bidder condition, where underbidding was most evident, is characterized by fairly mild perceived bonus incentives to lose but no perceived bonus incentives to win. Thus, even though the effects appear to be lower in the three-fourth condition,

they are consistently tilted in the direction towards losing, which appears to make the difference. In the English auctions three-bidder condition, segment 1 is the losing prone segment and segment 2 is a more indifferent segments.

Table 5. Mixture model estimates of the bidding functions.

Three Bidders SPSB			
Parameter	**Control**	**One-Fourth**	**Three-Fourth**
Segment 1–Alpha	0.40 **	0.52 $^{\pm}$	0.14 **
Segment 1–Beta	0.36 **	0.34 $^{\pm}$	0.00
Segment 2–Alpha	0	0.22 **	0.06 **
Segment 2–Beta	0.07 **	0.30 **	0.00
Segment 3–Alpha	0	0	0
Segment 3–Beta	0	0	0
Proportion Segment 1	0.74	0.63	0.23
Proportion Segment 2	0.14	0.21	0.28
Proportion Segment 3	0.12	0.16	0.49
Log-Likelihood	−1739	−2535	−1648
Six Bidders SPSB			
Parameter	**Control**	**One-Fourth**	**Three-Fourth**
Segment 1–Alpha	0.582 **	0	0
Segment 1–Beta	0	0.227 **	0.383 **
Segment 2–Alpha	0.166 **	0.132 **	0
Segment 2–Beta	0.192 **	0	0
Proportion Segment 1	0.874 **	0.673 **	0.652 **
Log-Likelihood	−1441	−1599	−1756
Three Bidders English–Only Pivotal Bids (Second Price) Used			
Parameter	**Control**	**One-Fourth**	**Three-Fourth**
Segment 1–Alpha	0.24 **	1.16 **	0.91 **
Segment 1–Beta	0.07	0.81 **	0.70 **
Segment 2–Alpha	−0.01	0	−0.02
Segment 2–Beta	−0.02	0	0
Proportion Segment 1	0.62 **	0.71 **	0.70 **
Log-Likelihood	−372	−1021	−583
Six Bidders English–Only Pivotal Bids (Second Price) Used			
Parameter	**Control**	**One-Fourth**	**Three-Fourth**
Segment 1–Alpha	0.087 **	0	0
Segment 1–Beta	0	0.158 **	0.623 **
Segment 2–Alpha	0.012 **	0.954 **	0
Segment 2–Beta	0	0.949 **	0.151 **
Proportion Segment 1	0.75 **	0.87 **	0.82 **
Log-Likelihood	−117	−521	−681

** Parameter significant at $p < 0.05$; $^{\pm}$ Alpha and beta are individually not significant at $p < 0.05$ but jointly significantly different from 0 ($p < 0.05$).

5. Conclusions

In this work, we study bidding behavior in auctions with price externality. Such auctions are understood in the literature to be the closest theoretical abstraction to charity auctions—auctions where the levels of proceeds presumably affect the utility of the bidder. Not all charity giving falls under this classification. For example, a bidder might donate to charity auctions for warm glow [29,30]—utility from the

act of giving that is decoupled from the amount raised. Therefore, in applying the insights from this work to charity auctions, one should take care to find auctions where the benefit to the bidder is relatively direct. Examples include school or church auctions, where the proceeds go towards enhancing services to members who are also the bidders in the auctions.

One of the variables that the investigation focused on is the *multiplier*. This multiplier serves as an abstraction for the degree that the proceeds directly affect the bidder. This has a direct parallel to the multiplier in the public good literature [31] and the implied marginal per capita return [32].

We find that while the multiplier—the direct benefit from the proceeds—can encourage bidders to increase bids, it can also result in ambivalence towards winning which might lower bid levels. Consistent with the latter, we observe significant underbidding in our controlled laboratory experiments. Bidders bid significantly below their valuations in most conditions and well below the theoretical optimal bid.

The empirical findings in the laboratory stand in sharp contrast to theoretical predictions, which suggest that bidders should bid more than their valuations and these bids should increase in the multiplier.

The pattern of underbidding includes observed bidders' reluctance to place bids in auctions with higher multipliers. That is, the patterns suggest that bidders facing a moderate to high multiplier do not bid aggressively. We propose that the observed pattern can be explained by bidders perceiving higher benefit to losing than to winning, as expressed by a higher alpha relative to beta in the theoretical model. The current results, including the estimation of the mixture model, suggest that the observed pattern is driven largely by a higher perceived benefit from losing than from winning.

The mixture model estimates rely on a particular utility specification. Clearly, there are alternative specifications for utility that might better fit the observed patterns. The most obvious modification would be to add concavity. This is not possible in the present heterogeneous estimation as it is not reliably separately identifiable by subpopulation from the other parameters. Estimating and correctly identifying more complex yet heterogeneous utility functions that will explain the observed patterns will require a different experimental design (e.g., varying the multiplier within subjects rather than between subjects). That said, the fact that subjects consistently bid below their valuations in the $n = 3$ conditions is pretty convincing evidence of loser incentives looming larger. This is because for any utility curvature, a bidder will not bid under their valuation in any of the conditions unless that incentive to lose is overweighted relative to the incentive to win. The pattern we uncovered is not invariant to incentives and competition. As the number of competing bidders increases from three to six, it changes both the total welfare stake and the competitive intensity.

In terms of social welfare, if we think of the auction with price externality as a social dilemma game and look at the extent of social inefficiency from underbidding [33], we can see that the 6-bidder settings have a greater extent of social inefficiency from low bids and this social incentive could be playing a role in reducing underbidding. In terms of competitive intensity, 6-bidder conditions are more likely to involve a bidding war that will bring bids up. Therefore, the 6-bidder auctions are characterized by far less underbidding.

This set of empirical results adds to a larger empirical literature indicating an unresolved puzzle regarding motivations in auctions with price externality. While empirical studies such as [20,21,34] report significant and respectable charity premiums, other studies such as [12] do not find a charity premium for more controlled settings.

The present set of results attempted to shed light on one possible explanation for this inconsistency—namely the asymmetric perceptions of the benefits from winning and losing. This asymmetry will presumably be particularly critical in settings that do not have endogenous participation like the setting here as well as the settings in [12] and less problematic in settings with bidder self-selection like the auctions in [20,21,34] where bidders choose to bid in auctions with price externality[8].

Moreover, we found that despite a substantial level of underbidding, as the multiplier increases winning prices significantly increase. This is an encouraging finding relative to [12], who find no significant revenue increase in sealed bid settings. It suggests that revenues can be increased, relative to no-externality settings, by conducting auctions with price externality with charitable bidders. Both the present work and [12]'s work suggest that the increase in revenues will be substantially less than the theoretical predictions, but our demonstrated increase in revenues is more optimistic than past findings.

In our auctions with price externality we see decreased competitive intensity as manifested by fewer bids. We attribute this behavior to the perceptions about benefits from winning and losing.

There are of course other variables that may influence competitive bidding behavior in charity auctions. For example, [21] observe reduced bidder entry in auctions with price externality relative to non-charity auctions. Other variables that may differentially influence competitive behavior in charity auctions include altruistic preferences [37] and "see and be seen" preferences [12]. That is, some

---

[8] Elsewhere in this literature, theoretical predictions do not seem to hold very well. For example, [35] report the results of four sealed bid auctions with price externality for four different preschools, comparing revenues of a first price, second price and all-pay auctions. In contrast to the theoretical work [35,36], they find that first price auctions revenues are highest followed by second price and finally all-pay auctions. They attribute this discrepancy to endogenous bidder participation.

individuals behave in a seemingly more altruistic manner when they think they are observed by others [38,39] and when they observe others behaving altruistically [40][9]. These issues merit further investigation.

Acknowledgments: This research was supported by grants from the Social Sciences and Humanities Research Council of Canada and the University of Alberta McCalla Professorship and GRA Rice Faculty Fellowship.

Author Contributions: Both authors contributed equally to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivation and Proofs

Let v denote one's valuation, b denotes his bid, B_S denotes the second highest price in the auction, and α denotes the portion of the winning bid that every bidder gets. From here on, the parameter α is referred to as the *multiplier*. The payoff function is specified as:

$$\pi(v, b) = (v - b)[if\ win] + \alpha B_S \tag{1}$$

We show the solution for a second price auction, which is arguably equivalent to an English clock auction, as argued by [13] and proved in [14]. This bid function could also be interpreted to deliver the highest price a bidder would be willing to stay in for in a non-clock English auction, although this may not be the case with non-incremental bids [41].

The payoff maximization problem for a bidder with valuation v_i in the second-price auction is specified below [13,14].

$$\begin{aligned} \max_b \pi(v_i, b) = \\ \int_{\underline{v}}^{b} [v_i - (1 - \alpha_i) B_S(x)] dF(x)^{n-1} dx \\ + \alpha_i B_S(b)(n-1)F(b)^{n-2}(1 - F(b)) \\ + \int_{b}^{\overline{v}} \alpha_i B_S(x)[(n-2)(n-1)F(x)^{n-3}(1 - F(x))f(x)]dx \end{aligned} \tag{2}$$

[9] This observation could imply predictions in either direction—depending on whether individuals are altruistic towards the charities or towards other bidders. It is reasonable to assume that when bidding in auctions with price externality auctions with price externality, the bidder's altruism towards the charity dominates.

The first term represents the utility that the bidder gets in the event that he wins the auction. It is equal to his valuation minus the price he pays, discounted by $(1-\alpha)$ because every bidder gets the multiplier, α, multiplied by the winning price. The price paid is equal to the second highest bid, denoted B_S. We integrate that over the possible prices that the winner might pay. Since the payment price is the second highest bid by an opponent, we integrate this term over the opponent's valuation.

The second term defines the utility that the bidder receives from placing the second highest bid as he will set the price and will thus get his share of that price. This is multiplied by the probability that his bid is the second highest bid and thus the pivotal bid. Note that this is the only instance of B_S in this equation that has one's own bid, denoted here by b, as a term. That is, one's own bid b only affects the price when it is the second highest, and thus the pivotal, bid.

The final term represents the utility from coming in lower than second. Bidders whose bids are lower than second will get α times the second highest bid. Note that the second highest bidder and lower price bidders are similar in that they all receive α times the second highest bid. The fact that utilities for the second highest bidders and lower price bidders are expressed by different terms is not due to different utility considerations. Rather, the difference comes from the second highest bid determining the price, whereas lower bidders do not. This means that the second highest bidder directly impacts his own utility with his bid, where lower price bidders do not impact their own utilities with their bids.

Maximizing the objective function of Equation (2) for the case of bidders adopting symmetric bid functions results in a general solution [14].

$$B_S(v_i) = v_i + \int_{v_i}^{\overline{v}} \left(\frac{(1-F(x))}{1-F(v_i)} \right)^{\frac{1}{\alpha_i}} dx \tag{3}$$

Applying this to the case of the uniform distribution on the range [50,100], used in this research, the bid function can be expressed as shown below.

$$B_S(v_i) = v_i + \frac{(100-v_i)\alpha_i}{1+\alpha_i} \tag{4}$$

A critical element of this solution is that the bid function for a second-price auction with price externality is independent of the number of bidders, n. Also notice that when α is zero the solution simplifies to v, the bidding rule for a second-price independent private value non-auction with price externality. Based on these result, we state critical properties. Using these properties, we then form the bases for our experimental tests.

An important feature of the present environment is that winning and losing bidders receive the same additional payoff, α_i, from each dollar of revenue to the charity. However, decision makers in the laboratory may or may not perceive the incentive to winners and the incentive to losers as being of the same magnitude. Following [12,13], we allow each bidder to possess two parameters α and β. [12] describe the parameters α and β as denoting the utility a bidder gains for each additional unit of revenue the auctioneer raises when the bidder loses the auction and when the bidder wins the auction, respectively. We refer to parameter α as the *incentive to losers* and to parameter β as the *incentive to winners.* The objective function is now changed to equation (2′):

$$\begin{aligned} \max_b \pi(v_i, b) = \\ \int_{\underline{v}}^{b} [v_i - (1-\beta_i)B_S(x)]dF(x)^{n-1}dx \\ +\alpha_i B_S(b)(n-1)F(b)^{n-2}(1-F(b)) \\ +\alpha_i \int_{b}^{\overline{v}} B_S(x)[(n-2)(n-1)F(x)^{n-3}(1-F(x))f(x)]dx \end{aligned} \tag{2′}$$

The optimal bid function gets restated in equation (4′):

$$B_S(v_i) = \left\{ \begin{array}{cc} \frac{v_i(1-\beta_i+\alpha_i)+100\alpha_i}{(1-\beta_i+2\alpha_i)(1-\beta_i+\alpha_i)} & if\ \alpha_i > 0 \\ \frac{v_i}{(1-\beta_i)} & if\ \alpha = 0 \end{array} \right\} \tag{4′}$$

Appendix B. Experiment Instructions

Experimenter: Welcome. This is a study in decision making. You will participate in several auctions, and if you pay careful attention and make wise decisions you will earn a considerable amount of cash which will be paid to you in private and in cash at the end of this study. Please carefully read the instructions. After this, we first run two practice auctions. Then you have an opportunity to ask questions before we start with the study. Please do not touch the computer until I instruct you to do so.

Instructions for Research Study

Welcome. This is a study in decision making. You will participate in several auctions, and if you pay careful attention and make wise decisions you will earn a considerable amount of cash which will be paid to you in private and in cash at the end of this study.

What to do in this study? The study will involve 10 separate rounds of auctions. In each round, you will place *bids* in tokens (Exchange rate 1 token = 2 cents) for a virtual item. These are "English" auctions where you keep on bidding, and prices keep going up, till only the winner remains.

How do I win? You *win* an auction if your *bid* is the highest for an item from among all bids submitted for this item. For example, if the highest bids submitted for an auction by bidders 1, 2, and 3 are 54, 48, and 87, respectively, then the number 87 is the highest and bidder 3 wins.

How can I tell if I am winning? The winning bid and bidder for each item will always be at the bottom of the list of bids for that item.

What do I win? When you win, you get your value for the item (shown in the top left corner of the screen), [minus 75% (three-fourth) of your winning bid]. To help you determine how much you will win, the CALCULATE button will compute the buyer net for any bid you enter and the computed amount will be displayed underneath the CALCULATE button as shown in the screen below.

What do I get when I lose? (When you lose, you get 25% (one fourth)) of the winning bid for the auction.

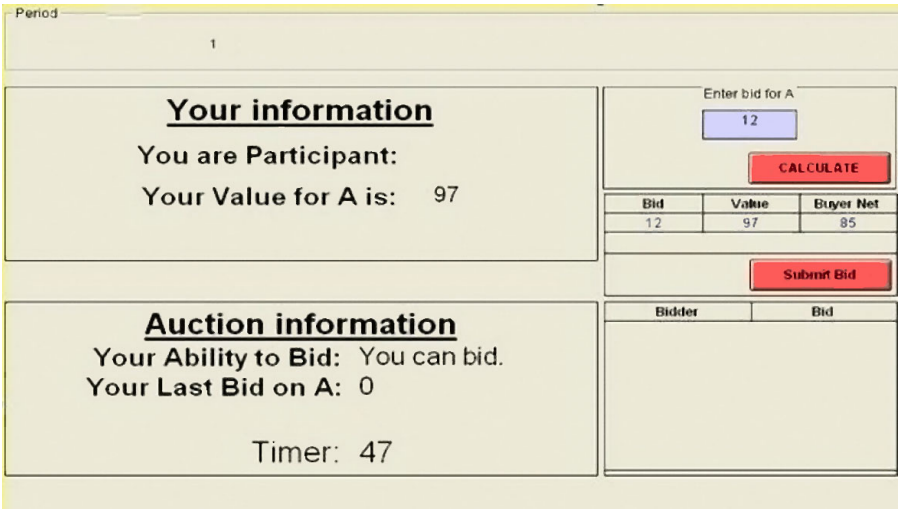
How to place a bid? To place a bid, you enter your bid in the box under "Enter a bid". Then press CALCULATE. Next, highlight the row underneath the CALCULATE button and press the Submit Bid button. The computer will then ask you to confirm one more time and you are done.

Can I place more than one bid? Yes, but not consecutively. You will need to wait for at least one bid by another bidder before you can enter a follow-up bid.

How much time do I have to place bids? You have **two minutes** to place bids. If anyone places a bid anytime in the last 15 s, the time will be **extended by 15 s**. So for example, if there are 3 s remaining and you place a bid, the timer will reset and there will be 15 s remaining from that point on.

What do I know about other bidders? (There are two other bidders competing with you.) Their bids will be shown on the screen under the Submit Bid button. You will not know their values. However, all values are randomly drawn from **50 to 100**. Each round a new set of numbers is independently generated for each participant.

We first start with 2 practice rounds after which you can ask questions.



References

1. Steitzer, S. Online School Auctions Curtail Hassles but Also the Socializing. Available online: http://www.wsj.com/articles/SB10208067829443850 80 (accessed on 12 February 2014).
2. Navarro, P. Why do corporations give to charity? *J. Bus.* **1988**, *61*, 65–93.
3. Varadarajan, P.R.; Menon, A. Cause-related marketing: A coalignment of marketing strategy and corporate philanthropy. *J. Market.* **1988**, *52*, 58–74.
4. Lichtenstein, D.R.; Drumwright, M.E.; Braig, B.M. The effect of corporate social responsibility on customer donations to corporate-supported nonprofits. *J. Market.* **2004**, *68*, 16–32.
5. Brown, T.J.; Dacin, P.A. The company and the product: Corporate associations and consumer product responses. *J. Market.* **1997**, *61*, 68–84.
6. Strahilevitz, M.; Myers, J.G. Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *J. Consum. Res.* **1998**, *24*, 434–446.
7. Pracejus, J.W.; Olsen, G.D.; Brown, N.R. On the prevalence and impact of vague quantifiers in the advertising of cause-related marketing (CRM). *J. Advert.* **2003**, *32*, 19–28.
8. Handelman, J.M.; Arnold, S.J. The role of marketing actions with a social dimension: Appeals to the institutional environment. *J. Market.* **1999**, *63*, 33–48.
9. Sen, S.; Bhattacharya, C.B. Does doing good always lead to doing better? Consumer reactions to corporate social responsibility. *J. Market. Res.* **2001**, *38*, 225–243.

10. Luo, X.; Bhattacharya, C.B. Corporate social responsibility, customer satisfaction, and market value. *J. Market.* **2006**, *70*, 1–18.
11. Engelbrecht-Wiggans, R. Auctions with price-proportional benefits to bidders. *Games Econ. Behav.* **1994**, *6*, 339–346.
12. Isaac, R.M.; Pevnitskaya, S.; Salmon, T.C. Do preferences for charitable giving help auctioneers? *Exp. Econ.* **2010**, *13*, 14–44.
13. Salmon, T.C.; Isaac, R.M. Revenue from the saints, the showoffs and the predators: Comparisons of auctions with price-preference values. *Res. Exp. Econ.* **2006**, *11*, 1–30.
14. Engers, M.; McManus, B. Auctions with price externality. *Int. Econ. Rev.* **2007**, *48*, 953–994.
15. Mak, V.; Rapoport, A.; Seale, D. Sequential search by groups with rank-dependent payoffs: An experimental study. *Org. Behav. Hum. Dec. Processes* **2014**, *124*, 256–267.
16. Popkowski Leszczyc, P.T.; Pracejus, J.W.; Shen, Y. Why more can be less: An inference-based explanation for hyper-subadditivity in bundle valuation. *Org. Behav. Hum. Dec. Processes* **2008**, *105*, 233–246.
17. Dholakia, U.M.; Basuroy, S.; Soltysinski, K. Auction or agent (or both)? A study of moderators of the herding bias in digital auctions. *Int. J. Res. Market.* **2002**, *19*, 115–130.
18. Ariely, D.; Simonson, I. Buying, bidding, playing, or competing? Value assessment and decision dynamics in online auctions. *J. Consum. Psychol.* **2003**, *13*, 113–123.
19. Heyman, J.E.; Orhun, Y.; Ariely, D. Auction fever: The effect of opponents and quasi-endowment on product valuations. *J. Interact. Market.* **2004**, *18*, 7–21.
20. Haruvy, E.; Popkowski Leszczyc, P.T. Bidder motives in cause-related auctions. *Int. J. Res. Market.* **2009**, *26*, 324–331.
21. Popkowski Leszczyc, T.L.; Rothkopf, M.H. Charitable motives and bidding in auctions with price externality. *Manag. Sci.* **2010**, *56*, 399–413.
22. Garcia, S.M.; Tor, A. The *N*-effect: More competitors, less competition. *Psychol. Sci.* **2009**, *20*, 871–877.
23. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
24. Zeithammer, R.; Adams, C. The sealed-bid abstraction in online auctions. *Market. Sci.* **2010**, *29*, 964–987.
25. Hortaçsu, A.; Nielsen, E.R. Commentary—Do bids equal values on eBay? *Market. Sci.* **2010**, *29*, 994–997.
26. Srinivasan, K.; Wang, X. Commentary: Bidders' experience and learning in online auctions: Issues and implications. *Market. Sci.* **2010**, *29*, 988–993.
27. Foreman, P.; Murnighan, J.K. Learning to avoid the winner's curse. *Org. Behav. Hum. Dec. Processes* **1996**, *67*, 170–180.
28. Ball, S.B.; Bazerman, M.H.; Carroll, J.S. An evaluation of learning in the bilateral winner's curse. *Org. Behav. Hum. Decis. Processes* **1991**, *48*, 1–22.
29. Andreoni, J. Giving with impure altruism: Applications to charity and Ricardian equivalence. *J. Polit. Econ.* **1989**, *97*, 1447–1458.

30. Andreoni, J. Impure altruism and donations to public goods: A theory of warm-glow giving. *Econ. J.* **1990**, *100*, 464–477.
31. Rapoport, A.; Eshed-Levy, D. Provision of step-level public goods: Effects of greed and fear of being gypped. *Org. Behav. Hum. Dec. Processes* **1989**, *44*, 325–344.
32. Isaac, R.M.; Walker, J.M. Group size effects in public goods provision: The voluntary contribution mechanism. *Quart. J. Econ.* **1988**, *53*, 179–200.
33. Mak, V.; Rapoport, A. The price of anarchy in social dilemmas: Traditional research paradigms and new network applications. *Org. Behav. Hum. Dec. Processes* **2013**, *120*, 142–153.
34. Elfenbein, D.W.; McManus, B. A greater price for a greater good? Evidence that consumers pay more for charity-linked products. *Am. Econ. J. Econ. Policy* **2010**, *2*, 28–60.
35. Carpenter, J.; Holmes, J.; Matthews, P.H. Auctions with price externality: A field experiment. *Econ. J.* **2008**, *118*, 92–113.
36. Goeree, J.K.; Maasland, E.; Onderstal, S.; Turner, J.L. How (not) to raise money. *J. Polit. Econ.* **2005**, *113*, 897–918.
37. Krishna, A.; Rajan, U. Cause marketing: Spillover effects of cause-related products in a product portfolio. *Manag. Sci.* **2009**, *55*, 1469–1485.
38. Glazer, A.; Konrad, K.A. A signaling explanation for charity. *Am. Econ. Rev.* **1996**, *10*, 1019–1028.
39. Romano, R.; Yildirim, H. Why charities announce donations: A positive perspective. *J. Public Econ.* **2001**, *81*, 423–447.
40. Frey, B.S.; Meier, S. Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *Am. Econ. Rev.* **2004**, *94*, 1717–1722.
41. Isaac, R.M.; Salmon, T.C.; Zillante, A. An experimental test of alternative models of bidding in ascending auctions. *Int. J. Game Theory* **2005**, *33*, 287–313.

Reciprocity in Labor Market Relationships: Evidence from an Experiment across High-Income OECD Countries

Israel Waichman, Ch'ng Kean Siang, Till Requate, Aric P. Shafran, Eva Camacho-Cuena, Yoshio Iida and Shosh Shahrabani

Abstract: We study differences in behavior across countries in a labor market context. To this end, we conducted a bilateral gift-exchange experiment comparing the behavior of subjects from five high-income OECD countries: Germany, Spain, Israel, Japan and the USA. We observe that in all countries, effort levels are increasing while rejection rates are decreasing in wage offers. However, we also find considerable differences in behavior across countries in both one-shot and repeated relationships, the most striking between Germany and Spain. We also discuss the influence of socio-economic indicators and the implications of our findings.

Reprinted from *Games*. Cite as: Waichman, I.; Siang, C.K.; Requate, T.; Shafran, A.P.; Camacho-Cuena, E.; Iida, Y.; Shahrabani, S. Reciprocity in Labor Market Relationships: Evidence from an Experiment across High-Income OECD Countries. *Games* **2015**, *6*, 473–494.

1. Introduction

There is now a large literature that models incomplete labor contracts between firms and their workers as a "gift exchange", where the worker perceives his/her wage as a "gift" and reciprocates with (costly) effort that benefits the firm. In the labor market context, this is referred to as the "efficiency wage hypothesis" (see [1,2]). Subsequently, a large literature of gift-exchange experiments has evolved (initially started by [3]; see the recent review by [4]), testing the efficiency wage hypothesis in a controlled environment. The typical findings of these gift-exchange studies are that both wage offers and effort levels are above the minimum and that repeated interaction between firms and workers leads to higher effort levels (e.g., [5–7]).

In the current study, we use a bilateral gift-exchange game as a work horse model to explore whether and to what extent behavioral differences are observed in a sequence of one-shot interactions (random-matching protocol) and also in repeated relationships (fixed matching) across subjects from five high-income OECD countries: Germany, Israel, Japan, Spain and the USA.

Although there is growing recognition that culture matters for a variety of economic outcomes (see, e.g., [8–11]), even models of social preferences that allow for individual-specific characteristics do not include an explicit culture- or

country-specific variable.[1] Moreover, despite the growing experimental literature comparing the behavior of subjects from different countries, there are only a few studies employing subjects from more than four countries (for example, [13–18]) and no such study of the labor market gift-exchange game. In the absence of a concrete model that predicts behavior across countries, our approach is similar to [16] (and also to other studies, such as [19,20]), in that we want to learn if there is evidence of differences in behavior across countries, rather than testing a particular theory.

Bilateral gift-exchange experiments have already been conducted in a variety of countries, mostly in Europe and the USA (see [4]). The contribution of this paper is to carefully control the experimental procedures across countries to allow for direct comparisons of results. In this respect, we utilize the inequity aversion model by [21] to study if differences in behavior could be explained by different aversions to advantageous inequity across countries. Moreover, our findings broaden the set of countries where comparable gift-exchange studies have been conducted, so that we can combine our results with previous research to investigate if some prominent socio-economic indicators could account for the differences in behavior.

Although there has not yet been a multi-country gift-exchange study, results from cross-country trust game experiments may shed some light on the expected results of bilateral gift-exchange studies due to the similarities in the two games. Like the gift-exchange game, the trust game ([22]) is a two-player game that measures reciprocity.[2] The static subgame perfect equilibrium under standard preferences is similar in both games (no reciprocity in the second stage and, hence, no giving in the first stage). Here, we briefly present the results of trust game studies involving at least two of our five countries.[3] Croson and Buchan [26] conducted a trust game experiment with subjects from China, South Korea, Japan and the United States, finding that women reciprocate more than men in all countries, but finding no differences across countries. Buchan *et al*. [27] conducted an additional trust game study in those four countries, finding that American and Chinese subjects are the most trusting, while Chinese and Korean subjects are the most reciprocal.[4]

---

1 In contrast to the *homo economicus* model assuming that individuals are rational money maximizers, models of "social preferences" assume that individuals do not only care about their own monetary gains, but also about other individuals (see a review by [12]).

2 In the trust game, a player can send a positive amount to another player. In the second stage, the amount is tripled, and the second player can send back a positive amount to the first player.

3 There are also a number of trust game studies comparing behavior between a pair of countries (e.g., [23–25]), but as these studies only include one of our samples of countries, we refrain from individually reporting their results.

4 Other studies that compare trust in Japan and the USA include survey evidence (e.g., [28]) and the findings from the one-shot prisoners' dilemma game by [29]. These studies find that Americans have a higher level of general trust than Japanese, but these findings may stem from differences in beliefs about the nature of social relationships rather than from differences in social preferences.

Hennig-Schmidt *et al*. [30] find, in a trust game study with German, Israeli, and Palestinian subjects, that Palestinians trust more than Germans, who trust more than Israelis. Back transfers by Germans and Israelis are not different, but both are lower than that of Palestinians. The results of trust games in different countries are summarized in the meta-analysis by [31], indicating no large differences in average behavior across our sample of countries, *i.e.*, subjects sent (sent back) between 51% and 59% (32% and 45%) of the possible amount (see Table A1 in the Appendix). As to the effect of repetition [32] compare the standard (one-shot) trust game with a repeated (fixed-matching) game. They observe that in the repeated game, subjects sent (sent back) 75% (56%) of the possible amount, while in the one-shot game, subjects sent (sent back) only 50% (38%) of the possible amount. In addition, Bohnet and Huck [33] study a binary trust game under stranger and partner matching. They find that 59% (61%) of the first (second) movers choose to trust (reciprocate) in partner matching compared to 32% (30%) in the stranger matching.

Note that there are two important differences between the bilateral gift-exchange game and the trust game: First, in contrast to the neutrally-framed trust game, our design imposes the labor market context (with which individuals in virtually all countries are familiar).[5] Second, the parameters in the bilateral gift exchange game are different from the trust game, and this may affect behavior. Particularly, in the bilateral gift-exchange game, a firm earns a very low payoff if its assigned worker does not reciprocate. By contrast, in the trust game, the first mover can assure himself/herself a significant payoff, regardless of the reciprocity of the second mover.

Our main findings can be summarized as follows: In the random matching treatments, we observe that Germans offer the highest wages, while Spanish offer the lowest. We do not observe differences in rejection rates across treatments. Further, we find that the efficiency wage hypothesis is confirmed in all countries, except for Spain, where at best, it is only weakly supported by the data. In terms of overall surplus, German subjects perform on average the best, while Spanish perform the worst. When the relationship between a firm and a worker is repeated (fixed matching treatments), we observe higher wages in all countries, except for Japan, and higher effort levels in all countries. This leads to higher surplus in repeated relationships compared to random matching treatments in all countries. German subjects also perform better than those of the other countries in terms of both effort and overall surplus in the repeated relationships.

---

[5] There is evidence that cooperation is sensitive to the particular framing (see [34] and the references therein). On the other hand, Fehr *et al*. [35] show, in a gift-exchange experiment, that formation in terms of "seller-buyer" instead of "firm-worker" does not matter much.

The paper is organized as follows: In the next section, we explain the experimental design and procedure used in our experiment. In Section 3, we present the results. In Section 4, we compare our results to similar gift-exchange experiments conducted in different countries. Finally, in Section 5, we summarize and conclude.

2. Experimental Design and Procedure

We use the bilateral gift-exchange (BGE) game initially used by [5,6,36] to compare the performance of undergraduate students from five high-income OECD countries: Germany, Israel, Japan, Spain and the USA.

At the beginning of the game, subjects are randomly assigned the roles of "firms" and "workers" (they retain these roles throughout the experiment). The experiment lasts 10 rounds. In each round, each firm is matched with one worker. At the beginning of each round, each firm receives an endowment of 120 ECU ("experimental currency units"). The firm, moving first, offers a wage (w) to the worker (between 20 and 120 ECU). Then, the worker chooses whether to accept or reject the offer. If the worker rejects, both the worker and the firm receive a payoff of 0. If the worker accepts the offer, he/she has to choose an effort level (e) from a finite grid between 0.1 and 1.0. Selecting an effort level above the minimum level is costly, as displayed in Table 1.

Table 1. Effort levels and effort costs.

Effort Level (e)	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
Cost per effort level ($C(e)$)	0	1	2	4	6	8	10	12	15	18

Additionally, if a worker accepts the offer, he or she has to pay a fixed cost of 20 ECU, which in the instructions is labeled as a travel cost. At the end of each round, the payoff is calculated. If a worker accepts the offer, the firm's payoff for the actual round is determined by:

$$\Pi_i^F\left(w_i, e_j\right) = (120 - w_i)\, e_j \tag{1}$$

and the worker's payoff is given by:

$$\Pi_j^W\left(w_i, e_j\right) = w_i - 20 - C\left(e_j\right) \tag{2}$$

where w_i denotes the wage offer of firm i, e_j denotes the effort level of the corresponding worker j and $C\left(e_j\right)$ is the cost of worker j's effort.

At the end of the experiment, subjects earn their accumulated payoffs from the 10 rounds. We chose these particular design parameters (*i.e.*, cost and payoff functions) because they have been used in several previous gift-exchange studies

conducted in Austria, France, Hungary, Malaysia, Portugal and the USA. In Section 4, we discuss the findings of these experiments. In the static subgame perfect equilibrium (SPE) of this game under standard *homo economicus* preferences, the firms offer a wage of 21 (or 20) units, anticipating that workers will choose the lowest positive effort level, *i.e.*, $e = 0.1$. The resulting payoffs are 9.9 (or 10) ECU to the firm and 1 (or 0) ECU to the worker. An alternative outcome is predicted by the "efficiency wage hypothesis," that workers reciprocate with respect to a high wage offer with high effort and that firms anticipate this reciprocity. This more optimistic outcome is also predicted by the inequity aversion model [21]. According to this model, a worker's utility depends positively on his/her payoff, but negatively on the difference between his/her payoff and the payoff of the firm. If such workers are offered a sufficiently high wage, they would choose a positive effort that reduces the difference in payoff between themselves and the firm. Although this aversion to payoff inequality is not a direct measure of reciprocity, it can be thought of as a proxy for it.

In each country, we conducted both random and fixed matching treatments. Random matching (RM) represents a sequence of one-shot interactions between a firm and a worker, whereas fixed matching (FM) implies a repeated relationship. The subgame perfect equilibrium is the same in both random and fixed matching. However, due to the possibility to build up reputation, repeated interactions between firms and workers may yield different results than a sequence of one-shot interactions, since a selfish worker would reciprocate in early rounds if he/she believes that it will pay off in terms of future wage offers (see [5]). We therefore expect to observe higher wage offers and higher reciprocity in the FM treatments than in the RM treatments.

A total of 428 undergraduate students (mostly economics or business majors) participated in our experiment, which was programmed and conducted using the z-Tree experimental software [37]. The experiment was conducted at the University of Kiel (Germany), University Jaume I, Castellón de la Plana (Spain), the Max Stern Yezreel Valley College, Emek Yezreel (Israel), California Polytechnic State University at San Luis Obispo (USA) and Kyoto Sangyo University (Japan).

Upon entering the computer lab, the subjects were given 10 min to read the instructions, which included a set of four questions to test whether they understood the experiment.[6] Then, we read the instructions aloud and showed how to calculate the answers to the four questions from the instructions. An experimenter from Germany was present during each of the other countries' sessions to ensure that

[6] Instructions were adapted from [5]. If subjects gave at least one incorrect answer, the experimenters individually explained to them how the payoffs were determined.

the same protocol was followed in the different countries.[7] The instructions were translated from English to the relevant language and then back to English by two different persons. The exchange rate between ECU and the local currencies was calculated to have a similar purchasing power in each country.

3. Results

Table 2 provides the first glance at the performance of the different subject pools, while Figure 1 illustrates the mean (and standard errors) of wage offers and effort levels in each treatment. For the non-parametric analysis, we exclude the first and last round (due to possible "start-game" effects stemming from unfamiliarity with the task in the first round and "end-game" effects in the last round of the fixed-matching treatments).[8] As an observation for both the random and fixed matching treatment, we are using the average performance per worker over Rounds 2–9 (*i.e.*, in both the RM and FM treatments, the number of independent observations is equal to number of firms or workers). The procedure is explained below.

Table 2. Average outcomes in the different treatments (Rounds 2–9). RM, random matching; FM, fixed matching.

Treatment	All Offers		Accepted Offers			Rejected Offers	
(Number of Subjects)	$\overline{w}$	Joint-Π	$\overline{w}$	$\overline{e}$	Π-ratio	$\overline{w}$	Rejections
Germany RM (36)	61.37	53.97	62.07	0.31	3.81	28.33	0.02
Germany FM (36)	67.93	66.09	67.97	0.63	1.73	67.14	0.04
Spain RM (54)	49.15	37.04	50.40	0.14	5.06	27.83	0.05
Spain FM (52)	60.57	51.40	63.67	0.32	4.00	34.36	0.10
Israel RM (34)	59.66	50.44	60.63	0.26	4.32	27.50	0.02
Israel FM (36)	67.98	58.54	70.49	0.45	5.25	44.71	0.09
USA RM (42)	53.01	42.28	55.09	0.23	4.31	30.07	0.08
USA FM (42)	62.21	54.77	64.10	0.37	3.21	35.27	0.06
Japan RM (50)	55.84	47.80	56.66	0.26	4.28	23.60	0.02
Japan FM (46)	57.67	56.61	58.70	0.49	1.69	27.33	0.03

The symbols "$\overline{w}$" and "$\overline{e}$" indicate average wage and effort, respectively. "Joint-Π" indicates the surplus from a relationship. It is equal to $\Pi^W + \Pi^F$. "Π-ratio" is equal to Π^W/Π^F, where Π^W (Π^F) denotes the worker's (firm's) payoff. In calculating the "Π-ratio", we omitted three observations where $\Pi^F = 0$, but not due to a rejection of offer (two cases in the Spain FM treatment and one case in the Israel RM treatment).

---

[7] Two of the authors located in Germany were present in all sessions conducted in Kiel, Germany. One of these authors traveled to Spain, Israel and Japan, and the other one to the USA.

[8] By and large, we do not observe a time trend when including Rounds 2–9. More precisely, we regressed the variables "wage offer" or "effort level" on "round" and on "round square" in each of the treatments. The only significant variable at $p \leq 0.05$ is in the Spain RM treatment, where "effort level" is negatively affected by the "round."

3.1. Behavior in a Sequence of One-Shot Interactions

We first examine the results in the RM treatments, in which workers are matched with different firms in every round, thus leaving no opportunity for building up reputation or applying dynamic strategies. We start by using a robust rank order test to pairwise compare across countries. The *p*-values of these comparisons are shown in Table B1 in the Appendix. When pairwise comparing wage offers across countries, we find that German subjects adopting the roles of firms offer the highest wages (significantly higher than their counterparts from Spain, the USA and Japan). Spanish subjects representing firms offer significantly lower wages than those from the other countries (except for the USA). In addition, we observe that hardly any firm offers the SPE wage of 20 or 21 (less than 2% in any of the subject pools). When such a wage offer is observed, the worker either rejects it or selects the minimum effort level.[9]

(a) Wage (random matching)

(b) Wage (fixed matching)

(c) Effort (random matching)

(d) Effort (fixed matching)

Figure 1. Mean wage and effort levels (with standard errors) for accepted offers in the random- and fixed-matching treatment over Rounds 2–9.

[9] We count 14 cases in Rounds 2–9 with SPE wage offers of 20 or 21 ECU. In 10 cases, the workers rejected the offers, and in four cases, they chose the minimum effort.

Next, we want to learn how wage offers affect rejection and effort levels. For this purpose, we use the Hurdle model specification (see [38]) to capture the worker's two-stage decision: whether or not to reject the wage offer (by a logit model) and, if not rejected, which effort level to choose (by a truncated linear regression).[10]

Table 3 indicates that, first, rejections depend (significantly) negatively on the wage offers for all subject pools, but we observe no significant difference in the "propensity to reject" across countries. Second, we observe that in all countries, effort levels depend (significantly) positively on the wage offers.[11] We find no difference in reciprocity (effort levels per wage offer) across countries, except for a lower reciprocity in Spain (the significance level is $p = 0.059$). This means that for an average wage offer of 60 ECU (in Round 5), the average effort in both Germany and the USA is equal to 0.23, but in Spain, it is only 0.13. For an average wage offer of 80 ECU, the average efforts in Germany and the USA are 0.35 and 0.33, but only 0.15 in Spain.

Table 3. Hurdle model estimation for the random-matching treatments.

	Rejection	**Effort Level**
Germany	3.677 (2.386)	−0.063 (0.130)
Spain	−0.031 (94.991)	0.144 (0.108)
Israel	−1.679 (3.240)	−0.001 (0.175)
Japan	1.827 (8.048)	0.034 (0.108)
Wage	−0.151 *** (0.051)	0.005 *** (0.002)
Wage × Germany	−0.130 (0.090)	0.001 (0.003)
Wage × Spain	−0.021 (3.804)	−0.004 * (0.002)
Wage × Israel	0.036 (0.107)	−0.000 (0.004)
Wage × Japan	−0.121 (0.393)	−0.000 (0.002)
Round dummies (2–10)	yes	yes
Constant	4.00 ** (1.439)	−0.017 (0.110)
Observations	1080	1030
Wald chi2$(k-1)$	314.94 ***	9223.03 ***
Pseudo R^2 / R^2	0.37	0.29

The rejection decision is estimated using a logit specification. The effort level decision is estimated using a truncated linear regression. Standard errors are in parentheses. Both specifications are estimated using bootstrapped clustered robust standard error (by session) with USA as the benchmark country (1000 replications per specification, 369 and 535 of them were completed in the logit and linear regression, respectively). Finally, *, ** and *** denote equal to or less than the 10%, 5% and 1% significance levels, respectively.

[10] In a public good game, the decision whether or not to contribute is implicit in the amount contributed to the public good. By contrast, in our gift-exchange game, the decision whether or not to reject an offer is explicit. Thus, hurdle models are especially suitable for the gift-exchange game.

[11] Using a joint-significance chi-squared test, we find that wage coefficients (*i.e.*, (Wage) + (Wage × Country)) are significantly different from zero ($p < 0.01$ in all countries).

Although workers reciprocate with higher effort when receiving higher wages, they do not attempt to split payoffs evenly. In fact, for no pair of matched subjects did the firm earn a positive payoff equal to the worker's payoff. For all matched pairs, workers' payoffs are found to be significantly larger than the firms' payoffs.[12] Even though on average the payoff ratios range between 3.81 (Germany) and 5.06 (Spain), we find no statistical difference between the subject pools.[13] Finally, in terms of surplus (measured by the joint payoffs of the firm and the worker), we find that Germans perform better than Spanish and U.S. Americans. Spanish subjects perform worse than subjects from all other countries, except for U.S. Americans. Our first result can now be formulated as follows:

Result 1 (random matching): In the treatments representing a sequence of one-shot relationships, we find that: (i) German subjects offer higher wages than the other subject pools (except for Israelis), while Spanish offer lower wages than the other pools (except for U.S. Americans); (ii) rejections and effort levels (per given wage offer) are not different across countries, except for Spanish subjects, who reciprocate with less effort (per wage) than their counterparts in the other countries; (iii) payoff inequality between workers and firms is not different across countries; and (iv) as to surplus, German subjects outperform Spanish and U.S. American subjects, while Spanish subjects fall behind all of the other countries' subjects, except for the U.S. Americans.

Models of social preferences are often used to explain the higher levels of effort observed in gift-exchange experiments compared to the standard economic prediction. Along these lines, we estimate country-specific parameters of the inequity aversion model of [21] and use these estimates to show how different norms regarding social preferences may explain observed differences across countries. As noted above, in the FM treatments, there are other reasons why workers may give high effort that are not based on social preferences, so we focus only on the RM treatments for this analysis. In their model, Fehr and Schmidt assume that individual utility depends on an individual's payoff, but is also negatively affected

---

[12] We used a median test to study if the payoff ratio is equal to one. We reject this hypotheses in all treatment at $p < 0.01$.

[13] Notably, the results suggest that in this game, the workers have more power than firms. This feature deviates from most labor markets with an excess supply of workers.

by the difference between the individual's payoff and the partner's payoff. Formally, the utility of worker j is:

$$U_j^W(e_j) = \Pi_j^W\left(w_i, e_j\right) - \alpha \max(\Pi_i^F\left(w_i, e_j\right) - \Pi_j^W\left(w_i, e_j\right), 0) - \beta \max(\Pi_j^W\left(w_i, e_j\right) - \Pi_i^F\left(w_i, e_j\right), 0) \quad (3)$$

where $\Pi_j^W\left(w_i, e_j\right)$ denotes the payoff of worker j, receiving a wage offer w_i from the respective firm and choosing an effort e_j. Similarly, $\Pi_i^F\left(w_i, e_j\right)$ represents the payoff of firm i offering a wage offer w_i and receiving an effort e_j from the respective worker.

The β and α parameters are the worker's marginal utility loss from advantageous and disadvantageous inequality, respectively. Fehr and Schmidt further assume that $0 < \beta < 1$ and $\alpha \geq \beta$. In the gift-exchange game, workers with a high β may choose a higher effort level to reduce the disutility from an unequal payoff. However, the α parameter will not affect a worker's choice in most situations (see [39]). The α parameter is defined so that it only affects a worker's choice when a worker is earning less than the firm. In most situations in the gift exchange game, a worker can prevent earning less than the firm by choosing a low enough effort.[14] In the range of (high) efforts where α does affect utility, a decrease in effort will both increase the worker's payoff and reduce payoff inequality. Therefore, the inequity aversion model predicts that the worker would always want to reduce effort and avoid earning less than the firm (for any value of α). Hence, in the following, we estimate only the "aversion to advantageous inequity" parameter, β.

In the context of the worker decision in a gift exchange game, the inequity aversion model of social preferences offers similar behavioral predictions to models based on reciprocity. With both models of social preferences, workers are inclined to respond to high wages with high effort, either to reduce payoff inequality or to reciprocate with respect to the generous wage offer. It is not our goal to determine whether high effort levels stem from inequity aversion or reciprocity; instead, we consider the β estimates to be informative about whether there are some kinds of social preferences that influence worker behavior in gift-exchange games and how those social preferences vary across countries.

To estimate β, we assume that workers choose the effort level with the highest utility where utility for each effort level is determined as described in Equation (3) plus an error term. We assume that the random error terms

[14] The exception to this is when the worker is offered a wage below 30. In this case, the firm will earn more than the worker for any effort chosen; however, by rejecting the offer, both the firm and worker earn zero. Thus, a worker with a high α might choose to reject, suffering the loss in wages from rejection in order to get a more equal payoff. Because we have very limited data where wages are this low, we do not estimate α, instead focusing on the β parameter.

for each effort level have an independent, identical extreme value distribution. This framework leads to multinomial logit choice probabilities. Let $E \equiv \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ denote the set of possible effort levels, and let $e_j \in E$ denote the effort chosen by worker j. Then, the probability that a worker chooses effort level e_j is:

$$Pr(e_j) = \frac{\exp U_j^W(e_j)}{\sum_{k \in E} \exp U_j^W(k)} \tag{4}$$

We estimate β by maximum likelihood and bootstrap standard errors. Estimates for each country are shown in Table 4. The baseline model shows a single estimate for each country. To allow for heterogeneity among subjects, we also estimate a finite mixture model with multiple types of subjects in each country. Each type has a distinct β, and the proportion of subjects (θ) of each type is estimated along with the β estimates. To determine how many types to use, we continued to add types as long as the addition results in an improvement to the Bayesian information criterion (BIC). The BIC is equal to $BIC = -2LL + klog(n)$, where LL is the log likelihood, k is the number of parameters and n is the number of observations.[15]

Table 4. Estimates of the aversion to advantageous inequity parameter.

Country	Parameter	Baseline Model	Mixture Model		
			Type 1	Type 2	Type3
Germany	β	0.24 *** (0.02)	0.26 *** (0.08)	0.10 (0.08)	
	θ		0.72 *** (0.20)		
Spain	β	0.02 (0.01)			
	θ				
Israel	β	0.20 *** (0.04)	0.48 ** (0.21)	0.30 (0.21)	0.04 (0.12)
	θ		0.11 (0.14)	0.30 (0.20)	
USA	β	0.19 *** (0.03)	0.31 *** (0.07)	0.18 *** (0.06)	0.00 (0.27)
	θ		0.16 (0.20)	0.51 *** (0.14)	
Japan	β	0.21 *** (0.02)	0.24 *** (0.08)	0.09 (0.08)	
	θ		0.57 *** (0.13)		

Estimates are from maximum likelihood estimation with bootstrapped standard errors using data from workers in the RM treatments only. In the baseline model, a single β is estimated for each country. For the mixture model, distinct β are estimated for each type, and θ is the prior probability that a subject is of that type. Instead of estimating θ for the last type, it is set such that probabilities sum to one. To determine the number of types to estimate for each country, we added types until the Bayesian information criterion did not decrease. Standard errors are in parentheses. *, ** and *** denote equal to or less than the 10%, 5% and 1% significance levels, respectively.

[15] The BIC is a criterion for model selection that makes an adjustment to the log likelihood to penalize models with larger numbers of parameters.

Looking at the baseline model, note that Spain, the country with the lowest effort levels, has a β close to zero and that is not statistically significant ($p = 0.12$), indicating that Spanish workers gain little utility from reciprocal behavior that reduces the inequality in earnings between themselves and their firm. All of the other countries, on the other hand, have β estimates that are close to 0.2 and are statistically significant at $p < 0.01$ (with Germany a little higher than the other three countries). These subjects do have some utility gain from reciprocating with respect to their firms. Notably, our β estimates in the four countries other than Spain are quite close to the original calibration by [21] that had a median value of 0.25.

The mixture model provides more insight into the distribution of β within each country. First, note that in Spain, the addition of a second type did not improve the BIC, so no results are reported for the mixture model. In particular, both types estimated for Spain had β close to zero and that was statistically insignificant, not very different than in the baseline model. In both Germany and Japan, the mixture model leads to two types, one with a β of about 0.25 and another with β close to 0.1 (and not statistically significant). In these countries, a large fraction of subjects are of the type with the higher β (0.72 in Germany, 0.57 in Japan). From this, it appears that between three quarters and one half of the subjects in Germany and Japan are averse to advantageous inequity, while the remaining subjects show little to no aversion. The mixture model leads to three types in Israel and the USA. The USA is similar to Germany and Japan with about two thirds of the subjects demonstrating significant aversion to advantageous inequity. Finally, in Israel, slightly more than half of the subjects are not averse to advantageous inequity, but the remaining subjects have greater aversion than in the other countries. In short, the mixture model demonstrates that all countries have a considerable proportion of subjects who exhibit little or no aversion to advantageous inequity. Further, the distribution of β varies somewhat, but not greatly among four of the countries, while Spain stands alone in having no subjects with significantly positive β.

3.2. The Effect of a Repeated Relationship

To examine behavior in repeated relationships (like most existing labor markets), we conducted fixed matching (FM) treatments where the same firm and worker are matched together in all 10 rounds. In the following, we are using the robust rank order test to pairwise compare (i) between the RM and FM treatments (a within-country comparison that isolates the effect of a repeated interaction between a firm and a worker) and (ii) across the FM treatments (a between-country comparison in the treatments with a repeated interaction between a firm and a worker). The *p*-values of these comparisons are shown in Tables B2 and B3 in the Appendix. When comparing the performance between the random matching (RM) and fixed matching (FM) treatments within each country and across subject pools, we find that wage

offers are significantly higher under FM than under RM in all countries, except for Japan (and in the FM treatments, we find no systematic differences in wage offers across countries in the pairwise tests). Further, we also observe that a repeated relationship increases effort levels in all countries. As for the difference across countries in the FM treatments, we observe that effort is the highest in Germany ($p < 0.01$ for all countries, except for Japan with $p = 0.06$).

Next, when inspecting the workers' payoffs, we find these to be significantly larger than the firms' payoffs in all treatments. A repetitive relationship reduces payoff inequality between workers and firms for German and Japanese subjects, but not for Spanish, Israeli and American subjects. Moreover, in the FM treatments, we find that payoff inequality is lower in Germany and Japan than in each of the other countries ($p < 0.01$ in all pairwise comparisons). Finally, for all subject pools, we observe that repeated relationships increases surplus (joint payoff of firms and workers). Germany has significantly higher surplus than all countries, expect for Israel. However, in FM treatments, the surpluses of Spanish subjects are not lower than those of Israelis, Americans and Japanese. We summarize these results below.

Result 2 (fixed matching): In the treatments with a repeated relationship between a firm and a worker, we find that: (i) wage offers do not systematically differ across countries; (ii) Germans choose the highest effort levels; (iii) the lowest payoff inequality between workers and firms is observed among German and Japanese subjects; and (iv) Germans obtain the highest surplus, but by and large we observe no differences in surplus among the other subject pools.

Result 3 (random *vs*. fixed matching): The effect of repeated relationship between a firm and a worker leads to: (i) higher wage offers for all subject pools, except for the Japanese; (ii) higher effort levels in all countries; (iii) a reduction in payoff inequality between workers and firms among German and Japanese subjects; and (iv) higher surplus for all subject pools.

4. Comparison to Previous Gift-Exchange Studies

As was mentioned earlier, we chose a design that had already been implemented in several countries. We now want to inspect our results in light of these previous studies. Table 5 shows the average wage ($\overline{w}$) and effort levels ($\overline{e}$) in those studies, only including treatments that: (i) use the bilateral version of the game (*i.e.*, one worker is randomly assigned to one firm); (ii) employ the same payoffs and effort cost schedules as in [5,6]; and (iii) only provide information to subjects regarding their own relationships (*i.e.*, subjects do not receive information about other firm-worker pairs). Notably, all of the previous treatments satisfying (i)–(iii) were conducted in OECD countries, except for one study conducted in Malaysia.

When inspecting the values in Table 5 together with our results (summarized in Table 2), we find the following: First, the low effort observed in our RM treatment in Castellón, Spain, and the high effort observed in the RM and FM treatments in Debrecen, Hungary, seem exceptional.[16] When omitting those two outliers from the comparison, we observe that in the RM treatments, the average wage offers (effort levels) in previous studies range between 53.51 and 63.41 ECU (0.24 and 0.33). In our sample, the average wage offers (effort levels) range between between 53.01 and 61.37 ECU (0.23 and 0.31). Hence, aside from the very high effort in Debrecen and the very low effort in Castellón, the average behavior does not seem to largely differ across countries and also between our sample and previous studies.[17]

Table 5. Performance in comparable previous bilateral gift-exchange game studies.

Study	Place	Subjects	$\overline{w}$	$\bar{e}$
Random matching				
[6]	Linz; Vienna	Soldiers	62.36	0.30
[36]	Linz	Soldiers	≈56–66	-
[43]	Debrecen	Students	60.60	0.53
[5]	Vienna	Students	≈57–63	-
[41]	California	Students	54.82	0.30
[42]	Ohio	Students	59.76	0.31
[40]	Lisbon	Students	≈60–70	≈0.40–0.55
[44]	Lyon	Students	53.51	0.24
[7]	Penang	Students	63.41	0.33
Fixed matching				
[6]	Linz; Vienna	Soldiers	64.12	0.51
[43]	Debrecen	Students	61.10	0.73
[5]	Vienna	Students	57.60	0.47
[7]	Penang	Students	58.42	0.43

In the studies by [41,42], workers could not reject the wage offer (it also seems that way in [40]). The mean wages and efforts in [6] are inferred from the regressions on p. 326 (with T = 5) [36,40], and the random matching treatment by [5] did not report on mean wage or effort; we thus denote the wage and effort in these treatments by the highest and lowest average wage and effort in Figure 2a (p. 340) in Fehr *et al.* [36], Figures 1 and 2 (p. 411) in Pereira *et al.* [40] and Figure 1 (p. 7) in Gächter & Falk [5]. Next, [42] used a maximum wage offer of 100, but without a fixed cost of 20. For comparability we therefore added 20 to the wage offers (we show here their standard no-table treatment). Finally, "-" indicates those cases where values are not reported.

In the FM treatments, average wage offers (effort levels) of previous studies range between 57.60 and 64.12 ECU (0.43 and 0.51). In our data, average wage offers

[16] In addition, Pereira *et al.* [40] (see Figures 1 and 2 in Pereira's work) also indicate a high average effort level in a study conducted in Lisbon, Portugal.

[17] Moreover, when comparing the behavior of U.S. American subjects, we observe that the wage offers in our experiment conducted in San Luis Obispo, California (53.01 ECU), are very close to the wage offers from Berkeley, California [41] (54.82 ECU), but both offers are lower than the average from Ohio State University [42] (59.76 ECU). Average effort, on the other hand, is lower in our experiment (0.23) than in Berkeley (0.30).

(effort levels) range between 57.67 and 67.98 ECU (0.32 and 0.49, excluding Germany with an effort level of 0.63). Hence, average behavior does not considerably differ across these subject pools.

Next, we inspect the possible correlations between socio-economic indicators and performance in those countries where similar BGE experiments have been conducted (*i.e.*, our sample of countries combined with the sample shown in Table 5 above). Table A1 in the Appendix presents the socio-economic indicators. When using a non-parametric Spearman correlation between each of these indicators and the average wage offers or effort levels per country, we do not find any obvious correlation between these country-specific indicators and behavior in the RM treatments.[18] However, in the FM treatments, the indicators of "norms of civic cooperation" (NCC, measuring the efficiency for which a society is solving collective action problems; see [45]) and the "long-term orientation index" (LTOI, defined as "fostering of virtues oriented towards future rewards", [46] p. 239) are positively correlated with effort (LTOI: Spearman coefficient 0.66, including eight countries, NCC: Spearman coefficient 0.65, including six countries).[19] Nevertheless, due to the low number of observations, the evidence is not statistically significant. We conclude that much further research is necessary to establish the link between socio-economic indicators and subjects' decisions in economic experiments.

5. Discussion and Conclusions

Despite the growing popularity in studying cross-culture behavioral differences, there have been only a handful of experiments with subjects from four or more countries and, until now, no cross-country labor market gift-exchange game study. The aim of this study is to learn about systematic differences in behavior across high-income OECD countries in a labor market experiment and also if these differences grow larger or smaller when the employer-employee relationship is repeated.

Our findings can be summarized as follows: In the random matching treatments, we observe that Germans offer the highest wages, while Spanish the lowest. We do not observe differences in rejection rates across treatments. Further, we find that the efficiency wage hypothesis is confirmed in all countries, expect for Spain, where at best, it is only weakly supported by the data. The overall surplus of German subjects

---

[18] We exclude from the analysis the three treatments where the mean wage and effort levels are not indicated (*i.e.*, the RM treatments by [5,36,40]) or those countries without the respective socio-economic indicators (e.g., the World Value Survey does not provide values for Israel).

[19] We have used the NCC measure used by [16], who found that NCC is positively correlated with the punishment of free riders and negatively correlated with anti-social punishment in a public good game experiment.

is, on average, the highest, while that of Spanish subjects is, on average, the lowest. We also observe that in comparison with random matching, fixed matching increases wages in all countries, except for Japan, and increases effort levels in all countries. This leads to higher surplus in all countries. Finally, we observe that German subjects also perform better than those of the other countries in terms of effort and overall surplus under fixed matching.

In addition, we use the data from the random matching treatments to model differences across countries in a particular way, by estimating the "aversion to advantageous inequity" parameter of the inequity aversion model [21] in each country. We do not observe large differences in aversion to advantageous inequity in Germany, Israel, Japan and the USA. However, this aversion is considerably lower in Spain. We also show that the distribution of inequity aversion across subjects within a country is similar across the four countries (except Spain) and that all five countries have a considerable proportion of subjects who do not exhibit aversion to advantageous inequity.

The fact that Germans offer higher wages and effort than their American counterparts is consistent with the gift-exchange game study by [47], who speculate that country-specific norms are behind the lower wages and effort levels of their American subjects in comparison with similar gift-exchange studies conducted by Fehr and colleagues in Europe.[20] In addition, in a variant of a trust game conducted with international PhD students, [48] find that subjects from Northern Europe earn substantially more than their Southern European counterparts. This result is in line with our most remarkable differences in behavior between Germans (Northern European) and Spanish (Southern European).

In addition, the low offers by Spanish subjects in the one-shot treatment are consistent with the findings from ultimatum game experiments (see the meta-analysis by [49]). An additional example of low transfers made by Spanish subjects in a trust game is provided by [50], who conducted a binary trust game experiment with subjects from Morocco, France and Spain. These authors observe that Spanish subjects are significantly less trustworthy than subjects from Morocco or France.[21] Further, in contrast to what could be conjectured from the ultimatum game study by [19], we do not find higher wage offers by American subjects than by Israeli and Japanese subjects, and rejection rates are also not lower in Israel and Japan than in the USA. In line with [26] and at odds with [27,29], we do not find considerable

---

[20] Charness *et al*. [42] show that the results by [47] may stem from their use of a comprehensive payoff table that is found to reduce wage offers and effort levels.

[21] Yet, in a study of gift-exchange markets, Brandts and Charness [51] confirm the efficiency wage hypothesis using student subjects from Barcelona.

differences in trust (*i.e.*, wage offers) and reciprocity (*i.e.*, effort levels) between Americans and Japanese.

In short, we find that, while some prior results from games involving reciprocity and trust are confirmed by our experiments, in other cases, the results do not extend to the gift-exchange setting. Our results demonstrate the value of conducting further cross-country studies that provide evidence on the consistency and robustness of previous findings.

Acknowledgments: Financial assistance was provided by Universiti Sains Malaysia (1001/PSosial/816117). We would like to thank Avi Weiss and Gianluca Grimalda for useful comments and remarks. Finally, we would like to thank the editor and three anonymous referees for their comments and suggestions.

Author Contributions: Israel Waichman, Ch'ng Kean Siang, and Till Requate designed the study. Israel Waichman, Ch'ng Kean Siang, Till Requate, Aric P. Shafran, Eva Camacho-Cuena, Yoshio Iida, and Shosh Shahrabani conducted the study. Israel Waichman and Aric P. Shafran analyzed the data. Israel Waichman, Till Requate, and Aric P. Shafran wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Appendix File: Cross-Country Socio-Economic Indicators

In this section, we report on several socio-economic indicators and formulate how they are expected to impact behavior in a bilateral gift-exchange game study. The indicators are displayed in Table A1 (in addition to the findings of the of the trust game meta-analysis by [31]).

We start by inspecting two cross-nation indicators suggested by [46]. These authors define the power distance index (PDI) as "the extent to which the less powerful members of the institutions and organizations within a country expect and accept that power is distributed unequally" ([46] p. 61). In particular, we expect higher PDI to be related to lower rejection rates and high effort levels. A second cross-nation dimension of possible relevance is the long-term orientation index (LTOI), defined as "fostering of virtues oriented towards future rewards—in particular, perseverance and thrift"([46] p. 239).[22] Thus, we expect higher LTOI to be positively related to high wage offers (or high effort levels) in the repeated treatments.

A key notion as to why subjects offer high wages in a gift-exchange game is the trust a player has in receiving a "gift" (high effort) back. To this end, we use Item V23 of [52] (Wave 5: 2005–2009), which measures the percentage of those who responded with "most people can be trusted" to the following question: "Generally speaking, would you say that most people can be trusted or that you need to be very

[22] The author (Minkov) used items from the World Value Survey (see below) satisfying three conditions ("thrift as a desirable trait for children", "national pride"and "importance of service to others"). The first and third conditions seems to be highly relevant to our game, while the second trait measures "self-enhancement".

careful in dealing with people?" We expect trust to be positively correlated with high wage offers.

Table A1. Socio-economic indicators.

	Trust Game		National Dimensions		WVS		WJP
	Sent	**Sent back**	**PDI**	**LTOI**	**Trust**	**NCC**	**RoLaw**
				Our sample			
Germany	0.51	0.44	35	83	0.33	7.80	0.81
Spain	-	-	57	48	0.19	7.67	0.68
Israel	0.59	0.45	13	38	-	-	-
USA	0.51	0.34	40	26	0.39	7.69	0.73
Japan	0.58	0.32	54	88	0.36	8.29	0.78
				Previous experiments			
Austria	0.62	0.38	11	60	-	-	0.82
France	0.43	0.33	68	63	0.18	7.07	0.74
Hungary	0.51	0.40	46	58	0.28	7.84	0.58
Malaysia	-	-	104	41	0.08	6.23	0.57
Portugal	-	-	63	30	-	-	0.70

In the trust game, "sent" and "sent back" are the average amount sent (in the first stage) and sent back (in the second stage) from the respective available amounts. PDI stands for "power distance index", and LTOI stands for "long-term orientation index" [46]. "Trust" measures the percentage of people responding to the item V23 in Wave 5 of the World Value Survey (WVS) in 2005–2009 by "most people can be trusted." The last two rows are following [16]: "NCC" denotes "norms of civic cooperation." It is the (rescaled) average of Items V198–V200 in Wave 5 of the WVS. "RoLaw" denotes the Rule of Law Index developed by the World Justice Project (we use the latest values from 2015). Finally, "-" indicates those cases where values are not reported.

Finally, following [16], we include two additional indicators, "norms of civic cooperation" (NCC) and "Rule of Law Index" (RoLaw).[23] NCC is defined by [45] as a measure for the efficiency for which a society is solving collective action problems. We follow [16], measuring NCC as the average of three World Value Survey items where participants are asked whether a particular behavior can be justified or not. The statements are (Item V198) "Claiming government benefits to which you are not entitled", (V199) "Avoiding a fare on public transport"and (V200) "Cheating on taxes if you have a chance." The higher the average, the stronger are civic norms in the respective country.[24] RoLaw [53] captures the perception of how the rule of law is enforced within the country. This in particular refers to constraints on government power, the absence of corruption, open government, fundamental rights, order and

[23] Herrmann *et al.* [16] conducted a public good game with punishment in 16 locations, finding that both NCC and RoLaw are negatively and significantly correlated with anti-social punishment (punishment of prosocial cooperators), and NCC is positively correlated with the punishment of free riders.

[24] We follow [16,45], scaling the answers from "always justifiable (1)" to "never justifiable (10),"instead of the opposite.

security, regulatory enforcement, civil justice, criminal justice and informal justice. We expect that these two indicators (high NCC and RoLaw) are related to high wage offers and high effort levels.

As already pointed out in Section 4, when using a Spearman correlation to link these socio-economic indicators to differences in behavior across countries we do not observe obvious correlations in the random matching treatments. However, the LTOI and NCC indicators are possibly correlated with high effort levels in the fixed matching treatments. Nevertheless, as the number of studies are small, these variables should be further explored in cross-country experiments and in meta-analyses of experiments.

Appendix B. Appendix File: Additional Tables and Figures

Table B1. Statistical differences across random-matching treatments.

Outcome	Average		Germany	Spain	Israel	USA	Japan
$\overline{W}$ (accepted offers)	56.66	Japan	0.09	0.09	0.27	0.63	-
$\bar{e}$ (accepted)	0.26		0.17	0.00	0.52	0.16	-
Π-ratio (accepted)	4.28		0.84	0.43	0.91	0.73	-
Joint-Π (all offers)	47.80		0.18	0.01	0.54	0.17	-
$\overline{W}$ (accepted offers)	55.09	USA	0.07	0.38	0.16	-	-
$\bar{e}$ (accepted)	0.23		0.00	0.00	0.65	-	-
Π-ratio (accepted)	4.31		0.49	0.63	0.98	-	-
Joint-Π (all offers)	42.28		0.01	0.32	0.21	-	-
$\overline{W}$ (accepted offers)	60.63	Israel	0.41	0.00	-	-	-
$\bar{e}$ (accepted)	0.26		0.12	0.01	-	-	-
Π-ratio (accepted)	4.32		0.95	0.53	-	-	-
Joint-Π (all offers)	50.44		0.21	0.00	-	-	-
$\overline{W}$ (accepted offers)	50.40	Spain	0.00	-	-	-	-
$\bar{e}$ (accepted)	0.14		0.00	-	-	-	-
Π-ratio (accepted)	5.06		0.29	-	-	-	-
Joint-Π (all offers)	37.04		0.00	-	-	-	-
$\overline{W}$ (accepted offers)	62.07	Germany	-	-	-	-	-
$\bar{e}$ (accepted)	0.31		-	-	-	-	-
Π-ratio (accepted)	3.81		-	-	-	-	-
Joint-Π (all offers)	53.97		-	-	-	-	-

p-values of pairwise comparisons of different outcomes, using a two-sided robust rank order test, in the different fixed-matching treatments. Unit of observation: average outcome per worker (averaged over Rounds 2–9). "Π-ratio" includes only cases where the firm's profit is larger than 0. Finally, "Average" denotes the average value of the outcome of the comparison. For instance, the average wage offer in Japan is equal to 56.66 experimental currency units (ECU) (see Table 2).

Table B2. Statistical differences across fixed-matching treatments.

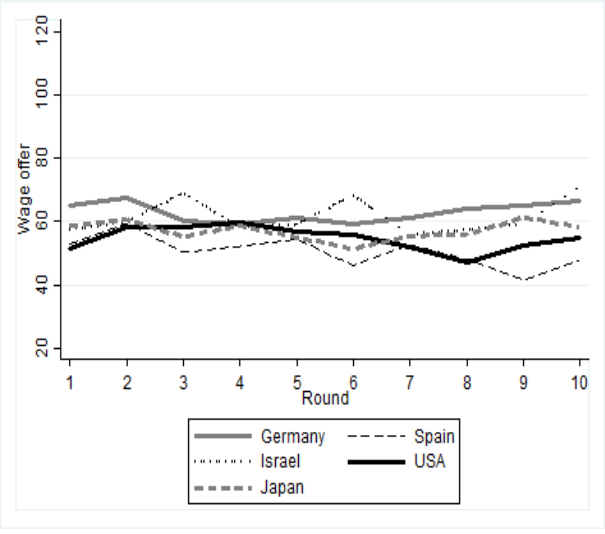
Outcome	Average		Germany	Spain	Israel	USA	Japan
$\overline{W}$ (accepted offers)	58.70	Japan	0.00	0.41	0.00	0.24	-
$\bar{e}$ (accepted)	0.49		0.06	0.00	0.42	0.13	-
Π-ratio (accepted)	1.69		0.72	0.00	0.00	0.00	-
Joint-Π (all offers)	56.61		0.06	0.24	0.70	0.57	-
$\overline{W}$ (accepted offers)	64.10	USA	0.09	0.73	0.09	-	-
$\bar{e}$ (accepted)	0.37		0.00	0.30	0.46	-	-
Π-ratio (accepted)	3.21		0.00	0.83	0.80	-	-
Joint-Π (all offers)	54.77		0.00	0.42	0.31	-	-
$\overline{W}$ (accepted offers)	70.47	Israel	0.67	0.19	-	-	-
$\bar{e}$ (accepted)	0.45		0.00	0.10	-	-	-
Π-ratio (accepted)	5.25		0.00	0.90	-	-	-
Joint-Π (all offers)	58.54		0.22	0.13	-	-	-
$\overline{W}$ (accepted offers)	63.67	Spain	0.22	-	-	-	-
$\bar{e}$ (accepted)	0.32		0.00	-	-	-	-
Π-ratio (accepted)	4.00		0.00	-	-	-	-
Joint-Π (all offers)	51.40		0.00	-	-	-	-
$\overline{W}$ (accepted offers)	67.97	Germany	-	-	-	-	-
$\bar{e}$ (accepted)	0.63		-	-	-	-	-
Π-ratio (accepted)	1.73		-	-	-	-	-
Joint-Π (all offers)	66.09		-	-	-	-	-

p-values of pairwise comparisons of different outcomes, using a two-sided robust rank order test, in the different fixed-matching treatments. Unit of observation: average outcome per worker (averaged over Rounds 2–9). "Π-ratio" includes only cases where the firm's profit is larger than 0. Finally, "Average" denotes the average outcome of comparison. For instance, the average wage offer in Japan is equal to 58.70 ECU (see Table 2).

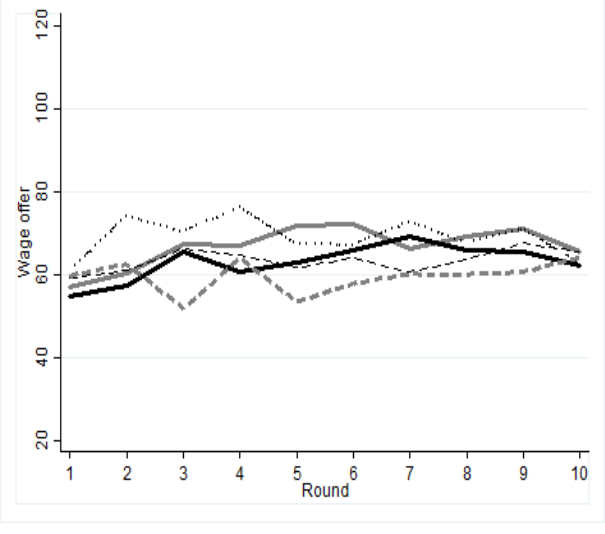
Table B3. The impact of repetition within each country.

Outcome	Germany	Spain	Israel	USA	Japan
$\overline{W}$ (accepted offers)	0.06	0.00	0.00	0.07	0.50
$\bar{e}$ (accepted)	0.00	0.00	0.00	0.00	0.00
Π-ratio (accepted)	0.00	0.18	0.63	0.19	0.00
Joint-Π (all offers)	0.00	0.00	0.05	0.00	0.06

p-values of pairwise comparisons of different outcomes, using a two-sided robust rank order test. Unit of observation: average outcome per worker (averaged over Rounds 2–9).

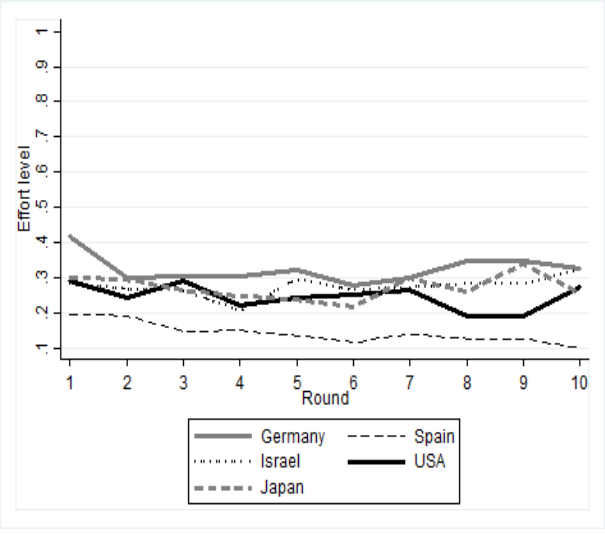


(a) Random-matching treatments

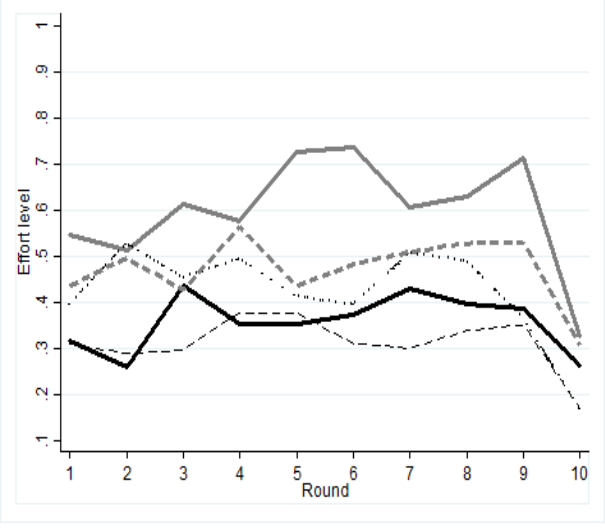


(b) Fixed-matching treatments

Figure B1. Evolution of average wages in the random- and fixed-matching treatments.



(a) Random-matching treatments



(b) Fixed-matching treatments

Figure B2. Evolution of average effort in the random- and fixed-matching treatments.

References

1. Akerlof, G.A. Labor contracts as partial gift exchange. *Q. J. Econ.* **1982**, *97*, 543–569.

2. Akerlof, G.A.; Yellen, J. The fair wage—Effort hypothesis and unemployment. *Q. J. Econ.* **1990**, *105*, 255–283.
3. Fehr, E.; Kirchsteiger, G.; Riedl, A. Does fairness prevent market clearing? An experimental investigation. *Q. J. Econ.* **1993**, *108*, 437–460.
4. Casoria, F.; Riedl, A. Experimental labor markets and policy considerations: Incomplete contracts and macroeconomic aspects. *J. Econ. Surv.* **2013**, *27*, 398–420.
5. Gächter, S.; Falk, A. Reputation and reciprocity: Consequences for the labour relation. *Scand. J. Econ.* **2002**, *104*, 1–26.
6. Kirchler, E.; Fehr, E.; Evans, R. Social exchange in the labor market: Reciprocity and trust versus egoistic money maximization. *J. Econ. Psychol.* **1996**, *17*, 313–341.
7. Siang, C.K.; Requate, T.; Waichman, I. On the role of social wage comparisons in gift-exchange experiments. *Econ. Lett.* **2011**, *112*, 75–78.
8. Alesina, A.; Giuliano, P. Culture and institutions. *J. Econ. Lit.* **2015**, in press.
9. Cardenas, J.C.; Carpenter, J. Behavioural development economics: Lessons from field labs in the developing world. *J. Dev. Stud.* **2008**, *44*, 311–338.
10. Fernández, R. Culture and economics. In *The New Palgrave Dictionary of Economics*; Durlauf, S.N., Blume, L.E., Eds.; Palgrave Macmillan: Basingstoke, UK, 2008.
11. Guiso, L.; Sapienza, P.; Zingales, L. Does culture affect economic outcomes? *J. Econ. Perspect.* **2006**, *20*, 23–48.
12. Fehr, E.; Schmidt, K.M. The economics of fairness, reciprocity and altruism—Experimental evidence and new theories. *Handb. Econ. Giv. Alt. Recipr.* **2006**, *1*, 615–691.
13. Bohnet, I.; Greig, F.; Herrmann, B.; Zeckhauser, R. Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *Am. Econ.Rev.* **2008**, *98*, 294–310.
14. Henrich, J.; Boyd, R.; Bowles, S.; Camerer, C.; Fehr, E.; Gintis, H.; Mcelreath, R. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* **2001**, *91*, 73–78.
15. Henrich, J.; McElreath, R.; Barr, A.; Ensminger, J.; Barrett, C.; Bolyanatz, A.; Cardenas, J.C.; Gurven, M.; Gwako, E.; Henrich, N.; *et al*. Costly punishment across human societies. *Science* **2006**, *312*, 1767–1770.
16. Herrmann, B.; Thöni, C.; Gächter, S. Antisocial punishment across societies. *Science* **2008**, *319*, 1362–1367.
17. Buchan, N.R.; Brewer, M.B.; Grimalda, G.; Wilson, R.K.; Fatas, E.; Foddy, M. Global social identity and global cooperation. *Psychol. Sci.* **2001**, *22*, 821–828.
18. Buchan, N.R.; Grimalda, G.; Wilson, R.; Brewer, M.; Fatas, E.; Foddy, M. Globalization and human cooperation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4138–4142.
19. Roth, A.E.; Prasnikar, V.; Okuno-Fujiwara, M.; Zamir, S. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *Am. Econ. Rev.* **1991**, *81*, 1068–1095.

20. Brandts, J.; Saijo, T.; Schram, A. How universal is behavior? a four country comparison of spite and cooperation in voluntary contribution mechanisms. *Public Choice* **2004**, *119*, 381–424.
21. Fehr, E.; Schmidt, K.M. A theory of fairness, competition and cooperation. *Q. J. Econ.* **1999**, *114*, 817–868.
22. Berg, J.; Dickhaut, J.; McCabe, K. Trust, reciprocity, and social history. *Games Econ. Behav.* **1995**, *10*, 122–142.
23. Willinger, M.; Keser, C.; Lohmann, C.; Usunier, J.-C. A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game. *J. Econ. Psychol.* **2003**, *24*, 447–466.
24. Buchan, N.; Croson, R. The boundaries of trust: Own and others' actions in the US and China. *J. Econ. Behav. Organ.* **2004**, *55*, 485–504.
25. Akaia, K.; Netzer, R.J. Trust and reciprocity among international groups: Experimental evidence from Austria and Japan. *J. Socio-Econ.* **2012**, *41*, 266–276.
26. Croson, R.; Buchan, N. Gender and culture: International experimental evidence from trust games. *Am. Econ. Rev.* **1999**, *89*, 386–391.
27. Buchan, N.R.; Croson, R.T.; Dawes, R.M. Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *Am. J. Sociol.* **2002**, *108*, 168–206.
28. Yamagishi, T.; Yamagishi, M. Trust and commitment in the United States and Japan. *Motiv. Emot.* **1994**, *18*, 129–166.
29. Hayashi, N.; Ostrom, E.; Walker, J.; Yamagishi, T. Reciprocity, trust, and the sense of control a cross-societal study. *Ration. Soc.* **1999**, *11*, 27–46.
30. Hennig-Schmidt, H.; Selten, R.; Walkowitz, G.; Winter, E.; Dakkak, I. Actions and Beliefs in a Trilateral Trust Game Involving Germans, Israelis and Palestinians. Presented at Experimental Science Association (ESA) Conference, Rome, Italy, 23 June 2007. ESA 2007 World Meeting at LUISS in Rome. Available online: http://static.luiss.it/esa2007/programme/papers/221.pdf (accessed on 29 Spetember 2015).
31. Johnson, N.D.; Mislin, A.A. Trust games: A meta-analysis. *J. Econ. Psychol.* **2011**, *32*, 865–889.
32. Cochard, F.; Nguyen-Van, P.; Willinger, M. Trusting behavior in a repeated investment game. *J. Econ. Behav. Organ.* **2004**, *55*, 31–44.
33. Bohnet, I.; Huck, S. Repetition and reputation: Implications for trust and trustworthiness when institutions change. *Am. Econ. Rev.* **2004**, *94*, 362–366.
34. Gerkey, D. Cooperation in context. *Curr. Anthropol.* **2013**, *54*, 144–176.
35. Fehr, E.; Klein, A.; Schmidt, K.M. Fairness and contract design. *Econometrica* **2007**, *75*, 121–154.
36. Fehr, E.; Kirchler, E.; Weichbold, A.; Gächter, S. When social norms overpower competition: Gift exchange in experimental labor markets. *J. Labor Econ.* **1998**, *16*, 324–351.

37. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178.
38. Nikiforakis, N. Punishment and counter-punishment in public good games: Can we really govern ourselves? *J. Public Econ.* **2008**, *92*, 91–112.
39. Charness, G.; Haruvy, E. Altruism, equity, and reciprocity in a gift-exchange experiment: An encompassing approach. *Games Econ. Behav.* **2002**, *40*, 203–231.
40. Pereira, P.T.; Silva, N.; e Silva, J.A. Positive and negative reciprocity in the labor market. *J. Econ. Behav. Organ.* **2006**, *59*, 406–422.
41. Charness, G. Attribution and reciprocity in an experimental labor market. *J. Labor Econ.* **2004**, *22*, 665–688.
42. Charness, G.; Frechette, G.R.; Kagel, J.H. How robust is laboratory gift exchange? *Exp. Econ.* **2004**, *7*, 189–205.
43. Falk, A.; Gächter, S.; Kovacs, J. Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *J. Econ. Psychol.* **1999**, *20*, 251–284.
44. Clark, A.E.; Masclet, D.; Villeval, M.-C. Effort and comparison income: Experimental and survey evidence. *Ind. Labor Relat. Rev.* **2010**, *63*, 407–426.
45. Knack, S.; Keefer, P. Does social capital have an economic payoff? A cross-country investigation. *Q. J. Econ.* **1997**, *112*, 1251–1288.
46. Hofstede, G.; Hofstede, G.J.; Minkov, M. *Cultures and Organizations: Software of the Mind*, 3rd ed.; McGraw-Hill: New York, NY, USA, 2010.
47. Hannan, R.L.; Kagel, J.H.; Moser, D.V. Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior. *J. Labor Econ.* **2002**, *20*, 923–951.
48. Bornhorst, F.; Ichino, A.; Kirchkamp, O.; Schlag, K.H.; Winter, E. Similarities and differences when building trust: The role of cultures. *Exp. Econ.* **2010**, *13*, 260–283.
49. Oosterbeek, H.; Sloof, R.; van de Kuilen, G. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Exp. Econ.* **2004**, *7*, 171–188.
50. Georgantzis, N.; Lacomba, J.A.; Lagos, F.; Milgram, J. *Trust and Reciprocity among Mediterranean Countries*; Economics Deprtament Working Paper; Economics Deprtament of Universidad Jaume I: Castellón de la Plana, Spain, 2013.
51. Brandts, J.; Charness, G. Do labour market conditions affect gift exchange? Some experimental evidence. *Econ. J.* **2004**, *114*, 684–708.
52. World Values Survey Association. *World Values Survey, Wave 5 2005–2008 Official Aggregate v.20140429*; World Values Survey Association: Madrid, Spain, 2014.
53. WJP Rule of Law Index 2015. Washington, D.C.: The World Justice Project. Available online: http://worldjusticeproject.org/rule-of-law-index (accessed on 8 September 2015).

Partner Selection and the Division of Surplus: Evidence from Ultimatum and Dictator Experiments

Priyodorshi Banerjee, Sujoy Chakravarty and Sanmitra Ghosh

Abstract: We study ultimatum and dictator environments with one-way, unenforceable pre-play communication from the proposer to the recipient, semantically framed as a promise. After observing this promise regarding how much the proposer will offer if selected, in our treatment conditions, recipients choose whether or not to select a particular proposer. We find that offers can increase in the ultimatum game both with non-competitive selection with a single potential proposer, and more so with competition, where the recipient chooses one of two potential proposers, as compared to the no selection baseline. Furthermore, the offer is rejected with higher probability if the promisemade by the selected proposer is higher than the eventual offer. Our dictator environment does not give the power to reject offers, thus selection power carries no benefits in the dictator game. Finally, independent of the game institution or proposer selection mechanism, promises provide credible signals for offers.

Reprinted from *Games*. Cite as: Banerjee, P.; Chakravarty, S.; Ghosh, S. Partner Selection and the Division of Surplus: Evidence from Ultimatum and Dictator Experiments. *Games* **2016**, *7*, 3.

1. Introduction

Analysis of bargaining usually proceeds on the assumption that a partnership has been formed. Sometimes, however, agents can choose whether to enter into a partnership, or whom to select as partner. In any bargaining situation, or productive enterprise where a partner is necessary, one would imagine that an economic agent would be careful about with whom to enter into partnership. This is especially so if the agent were the ex post weaker party. Yet the choice often has to be made on the basis of little more than non-binding assurances or promises regarding action consequent upon entering the partnership. Selecting a partner under such conditions may thus become an issue of concern. Indeed, determining what an agent believes when she hears promises, or how she selects upon hearing one or more promises, or whether such selection power is valuable, are of importance in understanding the partnership-formation process and the division of the resultant gains from trade.

For example, think of a homeowner wishing to engage a contractor to remodel her home. Many appear suitable given price, reputation and technical capacity. In

this situation, she may have to choose one on the basis of promises regarding the possibilities of delays, cost-overruns, and substandard work, knowing that these are not binding, and also that it is difficult to change contractors mid-way. Think of a supplier of intermediate goods, for example, who is thinking of entering into a relationship with a producer of final goods. There are sunk costs of reorienting production lines, making it difficult to change partnerships once one has been entered into. No matter how reputed the producer, the supplier will have to bear in mind at the time of deciding that promises regarding the likelihood of various hold-up problems that may arise in the future can be broken.

How does an agent decide in such situations, *i.e.*, what does she infer about intent on receiving a promise? Are partnerships preferred with those who promise more? Does the power to select a partner actually benefit the agent? Do such benefits, if any, depend on the mode of selection? What about promissors: do they display tendencies to keep promises once selected as partner?

In this paper, we report results from asymmetric bilateral bargaining experiments, using ultimatum and dictator games, designed to address these questions. Specifically, we analyze the impact of allowing the weaker party or recipient to select her partner or proposer on the division of surplus in the ultimatum and also the dictator game. Selection in our treatment conditions is done on the basis of pre-play communication: any potential proposer sends a promise about how much he will offer to the recipient if selected, *i.e.*, if given the right to propose. It is common knowledge that promises are not binding.

In our baseline ultimatum condition, there is a single potential proposer who sends a promise but faces no selection pressure, *i.e.*, the recipient is bound to award him the right to propose. Our two treatment conditions allow two different modes of selection. In the first, selection is non-competitive. There is a single potential proposer who sends a promise, and the recipient decides after obtaining the promise whether to select or refuse him, with the game ending and both parties getting nothing in the event of refusal. In the second, selection is competitive. There are two potential proposers each of whom sends a promise. The recipient decides after obtaining the promises which one of them to select.

Given that promises are not binding, the subgame perfect equilibrium outcome is invariant across the conditions and is the same as in the standard ultimatum game without communication: the (selected) proposer offers a minimal amount, and the recipient accepts. This leads to the hypothesis that introduction of selection power should make no difference to offers.

Alternate hypotheses may be entertained if players are behavioral. Since a proposer obtains the right to propose through a process of selection, this may endow him with ownership over the surplus, with the claim being larger if selection is competitive, since a comparison with another agent is involved. This leads to

the hypothesis that offers should be lower with selection power, especially with competitive selection.

On the other hand, the award of the right to propose arises through a decision on the part of the recipient, which she bases on communication received. This may lead her to use prior communication as a factor when deciding acceptance or rejection of the offer, which in turn can trigger better selection, and upward selection pressure on promises, especially in the competitive setting. This leads to the hypothesis that offers should be higher with selection power, especially with competitive selection.

Our data support the third hypothesis. We find that allowing the recipient to select her proposer increases offers made, with competitive selection having a larger impact. Specifically, compared to the baseline, non-competitive situation, selection increases offers by about 20 percent, with the difference weakly significant, while competitive selection increases offers by nearly 50 percent, with the difference strongly significant. Further, competitive selection increases offers by almost 25 percent compared to non-competitive selection, with the difference strongly significant.

Investigation of the decision to accept or reject offers yields a key to understand the channels through which communication-based selection power produces benefits for the recipient. We find in our treatment conditions that recipients use all information accrued on the path of play—offer as well as promise(s)—to determine their rejection decisions. Particularly, the likelihood of an offer being rejected is positively related to the promise made by the selected proposer, *ceteris paribus*. No such relationship is obtained however in the baseline condition where, in the absence of selection, promises hold little meaning. Overall, anticipation of recipients' rejection behavior causes proposers to issue credible promises, *i.e.*, make offers consonant with prior promise, in the treatments, but not in the baseline[1]. Together with recipients' power to select therefore, which creates upward pressure on promises, proposers are forced to make larger offers in the treatment conditions compared to the baseline, especially in the competitive setting. Thus, the impact of selection power is driven by the recipients' use at the final decision stage of information that is payoff-irrelevant at that stage but decision-relevant at an earlier stage.

The argument outlined above suggests that selection power can only deliver positive impact to a recipient when coupled with her power to reject offers. This implies that allowing promise-based selection in the dictator game, which is the ultimatum game with the recipient stripped of rejection power, should not have any

---

[1] This suggests that recipients should select higher promises with greater likelihood, *i.e.*, be credulous. We indeed found such a relationship in the competitive treatment but not in the non-competitive treatment, where the paucity of refusals prevented us from estimating a selection function.

discernible impact on offers made. Further, it should not lead to credible promises or credulity on the part of recipients. To confirm this intuition, we conducted three dictator conditions, one corresponding to each ultimatum condition. We found that promises were indeed not credible in the treatment conditions, higher promises were not necessarily selected with higher probability, and also that selection power did not lead to larger offers compared to the baseline.

The rest of the paper is organized as follows. We discuss related literature in the next section. Section 3 describes our treatments and procedure in detail, and lays out our empirical methodology and hypotheses. Section 4 contains results. Section 5 concludes. Appendix A and Appendix B, respectively, contain instruction and response sheets.

2. Related Literature

A large amount of literature exists on ultimatum and dictator games. Our paper is linked to three strands of the literature. To our knowledge, however, no prior paper has allowed the recipient to select whether she wishes to play or who becomes her proposer.

Whether a potential proposer actually obtains the right to propose is endogenously determined in our treatment conditions. Hence, our paper is connected to the literature studying whether how a player becomes the proposer affects outcomes: see, e.g., Hoffman, McCabe, Shachat and Smith [1]. The central difference is that the right is awarded through recipient decision, which is based on communication, in our environment, while prior literature has not explored direct selection power on the part of the recipient, with or without communication. It has not therefore explored how selection, offer and rejection may simultaneously depend on the content of prior communication. The closest paper in this strand is Navarro and Veszteg [2]. They study two-player interactions where an ultimatum game is preceded by a bidding stage, with the higher bidder becoming the proposer and the lower the recipient. They find that such bidding benefits the recipient at the expense of the proposer. Apart from the recipient having no selection power in their design, an additional difference is that the equilibrium outcome is the same in all our conditions, while the outcome in their treatment conditions is different from that in their control condition[2].

The proposer(s) in our treatment conditions are subjected to selection pressure. This connects our essay with the literature investigating whether outcomes are affected by the recipient facing multiple proposals (which changes the equilibrium

[2] Navarro and Veszteg [2] also find, supporting the conclusions of Gale, Binmore and Samuelson [3], that there is considerable learning, with the process not necessarily leading to a Nash equilibrium.

outcome): see, e.g., Roth, Prasnikar, Okuno-Fujiwara and Zamir [4] and Cox [5]. The key difference is that our recipients face at most a single proposal, but they may face one or more potential proposers, at most one of whom will obtain the right to propose. Further, the equilibrium outcome in our case does not vary across conditions.

Since non-binding communication is present in all our conditions, and is used as the basis for selection in our treatment conditions, our article is linked to the emerging literature analyzing the impact of communication in asymmetric bargaining games[3]. To our knowledge, the effect of promise-communication has not been discussed in ultimatum environments[4]. There is also a paucity of work on the impact of promise-communication in dictator environments. An exception is Vanberg [8], who finds dictators may have a preference for truth-telling.

Our paper is also connected to the literature studying partnership formation. Despite the centralityof partnership activity, partnership formation has received relatively limited experimental investigation. Prior papers have usually studied partner selection based on the history of play, with players interacting repeatedly, and the experimenter making the initial match: see for example Tullock [9], Coricelli, Fehr and Fellner [10] and Brandts, Cooper and Weber [11]. None of these have allowed for communication. Our article is closer to the nascent line scrutinizing partnership formation and performancein situations where partner selection is on the basis of non-binding communication: see Dulleck, Kerschbamer and Sutter [12], Beck, Kerschbamer, Qiu and Sutter [13], both on credence goods, and Goeree and Zhang [14], who examine principal-agent interactions in hidden information environments[5].

3. Design, Procedure, Methodology and Hypotheses

We first describe our experimental conditions in detail. The two control conditions are Dictator Baseline (DB) and Ultimatum Baseline (UB). DB is a standard dictator game (Forsythe, Horowitz, Savin and Sefton [15]) augmented with a prior promise stage. There is a proposer and a recipient, and a divisible surplus of 100 units. In the first stage, the proposer sends a (non-binding) promise about how much he will actually offer to the recipient out of 100 (a message of the form: I

---

[3] (It is additionally affiliated to the limited research on how choices in promise-communication conditions depend on the presence of punishment options: see, for example, Bochet and Putterman [6], who examine public goods problems. No paper we are aware of in this strand examines bargaining problems.)

[4] Although the game used by Ellingsen and Johannesson [7] is a trust game extended to include an ultimatum game element.

[5] Goeree and Zhang [14] find, in their communication treatments (2CNDR, 2CWDR and 3CWDR), that introducing competition, where the principal faces two agents, rather than one, lowers employment, efficiency and credibility of agent choice (see Figures 2 and 3, and Tables 2 and 3).

promise to give you x, with x belonging to {0,1,...,100}). The second stage then ensues where the proposer makes an offer (y belonging to {0,1,...,100}). The recipient is passive throughout and has to accept the offer. She gets the offered amount, and the proposer keeps the difference. Similarly, UB is a standard ultimatum game (Güth, Schmittberger and Schwarze [16]) augmented with a prior promise stage. Once again, there is a proposer and a recipient and a surplus of size 100. The proposer sends a promise in the first stage about how much he will actually offer, and then makes an offer in the second stage. The recipient chooses whether to accept or reject the offer in the third stage. If the offer is accepted, the recipient gets the offered amount and the proposer keeps the difference, while, if it is rejected, both get 0.

Our first set of treatment conditions are DS and US (S for selection). These are identical to DB and UB respectively, except that there is another stage, the selection stage, in between the promise and offer stages. Specifically, in DS, after hearing the promise (a message of the form: I promise to give you x if you select me), the recipient can decide to refuse the proposer, in which case the game ends and both parties get 0. However, if she selects the proposer, the game moves to a third stage where the proposer makes a binding offer. Similarly, in US, the recipient decides, after hearing the promise, whether to select or refuse the proposer. The game ends if there is refusal, with both parties getting 0, while it moves to the offer stage if there is selection.

Our second set of treatment conditions are DC and UC (C for competition). These are similar to DS and US respectively, except there are three players, two (potential) proposers and one recipient, and the game necessarily moves to the offer stage. In either environment, both potential proposers send promises in the first stage. The recipient then chooses one of them as the actual proposer or partner in the selection stage, whereupon the offer stage ensues, with the chosen proposer offering a division of the surplus.

3.1. Procedure

The choice in the first stage in all our conditions is a number interpreted as a promise. Any promise is transmitted to the recipient embedded in a message with a fixed semantic form (using the term "promise")[6]. Prior research shows that interactive, multi-round, or free-form communication is usually more effective in promoting positive outcomes in laboratory environments compared to communication which is unilateral, single-instance, fixed-form and numerical. We

[6] The term promise was given semantic focus because promises are distinct and familiar. Additionally, they seem to be somewhat special: evidence suggests that if free-form communication is allowed, messages regularly get coded explicitly as promises (see, for example, Ellingsen and Johanesson [7]).

chose the latter mode for our experimental procedure precisely for this purpose, as we had an interest in examining credibility (the tendency to make an offer consonant with prior promise) and credulity (the tendency to believe that a higher promise will yield a higher offer)in situations unlikely to promote them.

A single play of DB yields as data a promise in the first stage, and an offer in the second. A single play of DS yields a promise in the first stage, a selection choice in the second, and, if there is selection, an offer in the third. A single play of DC yields two promises in the first stage, a selection choice in the second, and an offer in the third. Single plays of UB, US and UC yield the same data as DB, DS and DC respectively, except for an additional acceptance/rejection decision in the final stage.

We ran one session for each of our conditions. Each session was hand run with a maximum time limit of two hours. Subjects were first assembled together, and instructions were read out[7]. After instructions were repeated and doubts clarified, subjects were split up. For DC and UC, one-third were assigned the role of recipients randomly, and moved to another room. After this, subjects were allotted identification numbers, and the experiment commenced. Roles remained fixed for the duration of the experiment. The other conditions proceeded similarly, except that half were assigned the role of recipients, and the other half proposers. In all conditions, subjects were given a reward based on payoff received. Subjects received between INR 100 and 500, a reasonable rate[8]. Payment was made at the end of the session in private.

We ran each condition for a few rounds to increase the amount of data collected in order to facilitate statistical analysis. For DC and UC, at the beginning of each round, the experimenter randomly matched each recipient subject ID with two proposer subject IDs, creating as many anonymous 3-member groups consisting of one recipient and two proposers, as the number of recipients[9]. This stranger matching was dissolved at the end of the round. The other conditions proceeded similarly, except that every matching was between one proposer subject and one recipient subject.

All sessions were run at Jadavpur University in Kolkata. Table 1 presents the number of rounds and subjects per condition.

---

7 Instructions and response sheets can respectively be found in Appendix A and Appendix B.

8 The purchasing power parity exchange rate between the Indian Rupee and the US Dollar for 2009 was 15 rupees to a dollar according to the Penn World Tables (Heston, Summers and Aten [17]).

9 Every proposer subject was matched to a single recipient subject.

Table 1. Session/condition details.

Session/Condition	Number of Proposer Subjects	Number of Recipient Subjects	Number of Rounds
UB	16	16	6
DB	16	16	6
US	17	17	6
DS	18	18	6
UC	60	30	4
DC	50	25	5

3.2. Empirical Methodology

The unit of observation in our analysis is the individual decision-maker. Since every subject made decisions over multiple rounds, our observations do not satisfy independence. We therefore adopt a cluster-adjusted regression approach uniformly throughout the paper, with clustering on the individual. Clustering offers a correction for deflation of standard errors which could result in overstated levels of significance. All regressions reported in this paper also use gender and round dummies as additional regressors. However, we do not report coefficients on these variables, as these were rarely significant.

3.3. Hypotheses

We lay out our four main null hypotheses for the ultimatum and dictator game experiments. Using game theoretic reasoning, we should not expect to see any impact of promise or competitive selection on offers. Following the subgame perfect Nash equilibrium, offers in our dictator treatments and offers and probability of acceptance in our ultimatum treatments should not diverge from the baseline for these games.

Our first null hypothesis (H1) is that selection power, with or without competition, has no effect on offers made, *i.e.*,—UC = US = UB in terms of average offer. Similarly, for the dictator game, average offers made across conditions should not be statistically different.

Our second hypothesis (H2) relates to the notion of credibility. We conjecture that proposers will not issue credible promises (observed offers will not be positively related to promise) in the treatments or the baseline for either the ultimatum game or the dictator game.

Our third hypothesis (H3) relates to the notion of credulity. We conjecture that non-credulous recipients will not believe that higher promises will yield enhanced offers, and not select these with greater probability in either the ultimatum or the dictator environment.

Our fourth hypothesis (H4) is connected to how the recipient decides whether to accept or reject an offer in the ultimatum game. We conjecture that recipients in our

selection treatments do not ceteris paribus take into account the promise associated with it, when deciding whether or not to reject an offer.

4. Results

Our results mainly focus on the ultimatum game, using the dictator conditions as controls. Section 4.1 explores H1 (treatment effects) and H2 (credibility) and compares our findings with those that are relevant in the extant literature. Section 4.2 examines H3 (credulity) and studies how recipients select promissors. Finally, Section 4.3 explores H4, *i.e.*,—the acceptance decision made by the recipient in the second stage of the ultimatum game.

4.1. Promises, Offers and Credibility

Findings from two surveys (Bearden [18], a large-scale review and Oosterbeek, Sloof and van de Kuilen [19], a meta-analysis based review of 37 papers) indicate that mean offer in the ultimatum game is most often between 40 and 50, and modal offer is 50. Means of offer and promise from both our ultimatum and our dictator conditions are shown in Table 2. The modal offer (not reported in Table 2) in all ultimatum conditions are 50, and mean offer for US is within this range. The mean offer for UC is just above the upper edge of this range, while mean offer for UB is somewhat lower than the lower edge of the range.

For our dictator experiments, our DB subjects promised to give about 31 percent on average to recipients, whereas the actual average amount given in DB is 16.5 which is well below the give rate of 27 percent found in a meta-study of dictator games by Engel [20] that references 616 treatments from 129 published papers.

Table 2. Mean offers and promises by treatment.

Treatment	Mean Offer	Mean Promise
UB	35.7 (96)	50.3 (96)
US †	42.9 (42.1) (100)	50 (102)
UC	52.6 (120)	68.9 (240)
DB	16.5 (96)	31.1 (96)
DS **	20.8 (19.5) (101)	51.4 (108)
DC	18.4 (125)	60.1 (250)

Total number of observations in parentheses; † For US, two promises were refused. Average offer and promise are reported from the set of selected promises. The mean for all 102 observations is given in square-braces; ** For DS, seven promises were refused. Summary statistics are reported from the set of selected promises. The mean for all 108 observations is given in square-braces.

To test H1, we need to compare averages across experimental conditions. Instead of comparing the simple averages given in Table 2 that are constructed from several rounds of the treatment and baseline conditions, we regress offers and promises on (i) game-treatment interaction dummies (with DB as the base category); (ii) gender

dummy and (iii) round dummies (these regressions are not reported here for the sake of brevity). This allows us to compare the conditional means of promises and offers between the different treatment conditions using a Wald test. For the ultimatum conditions, mean offer in US is greater from that in UB with marginal significance (F-test p-value = 0.0875). Mean offers for UC is significantly greater than that in UB (F-test p-value = 0.0003). Furthermore, mean offers in UC exceeds that in US (F-test p-value = 0.0034). Thus, selection power, especially with competitive selection, gave the recipient an advantage in the ultimatum *vis-à-vis* the dictator environment.

Adding in promise as a regressor along with treatment, gender and round dummies, results in the former explaining a significant amount of the variation in offers. This is presented in Table 3 below.

Table 3. Offer regression.

Variable	Dep. Var. = Offer
promise	0.27*** (0.08)
UB	14.55*** (4.85)
DS	−1.09 (5.58)
US	21.92*** (4.33)
DC	−5.88 (5.35)
UC	25.02*** (6.16)
constant	3.07 (3.19)
R^2	0.44
No. of obs.	638

Standard errors are in parentheses. "*", "**" and "***" respectively denote significance at the 10%, 5% and 1% levels.

Now, the mean offers in the treatments (US and UC) differ from that at baseline (UB) only at the 10 percent level of significance (US *vs.* UB, F-test p-value = 0.0923, UC *vs.* UB, F-test p-value =0.0645). Moreover, mean offer in UC is not statistically distinguishable from the mean offer in US(F-test p-value = 0.4025).

In the dictator game, we found mean offers for the three conditions to be statistically indistinguishable, as hypothesized. Thus, selection power gave the recipient no advantage in the dictator environment. Thus, H1 is rejected in the ultimatum environment but cannot be rejected in the dictator environment. Thus, selection power can only deliver positive impact to a recipient when coupled with her power to reject offers. Furthermore, the dictator conditions, which manipulate the method of promise selection for recipients but give no ability to recipients to reject unfavourable offers, provide control conditions to test the effect of the presence of this veto power. Wald tests for conditional mean offers, finds the ultimatum conditions

(UB, US, UC) to have significantly higher offers as compared to the corresponding dictator conditions (DB, DS, DC) at the one percent level[10].

For average promises in ultimatum games, the Wald tests indicate that UC is greater than UB (F-test p-value = 0.0008), US is not statistically different from UB (F-test p-value = 0.9773), and UC is greater than US (F-test p-value = 0.0000). For the dictator treatments, average promise in DC is greater than in DB (F-test p-value = 0.0000), DS is greater than DB (F-test p-value = 0.0056) and DC is marginally significantly greater than DS (F-test p-value = 0.0996). Thus, competitive selection generates an upward pressure on promises. Wald tests for conditional mean promises show that the average promises in UB and UC are higher than in the corresponding DB and DC conditions at the one percent level, though the average promised amount in US is not significantly different from DS. The most important Wald test comparisons of average offers and promises from the ultimatum and dictator treatments and have been summarized in Table 4. In summary, competitive selection significantly increases promises made in both the dictator and the ultimatum environments, but offers are somewhat higher with competitive selection only in the ultimatum environment. Furthermore, the presence of veto power (ultimatum game) generates significantly higher promises and offers *vis-à-vis* a situation with no veto power (dictator game).

Table 4. Wald Test results for ultimatum and dictator treatments.

Comparison	Ultimatum	Comparison	Ultimatum *vs.* Dictator
US *vs.* UB	Promise (US) $\approx$ Promise (UB) † Offer (US) > Offer (UB) *	US *vs.* DS	Promise (US) $\approx$ Promise (DS) Offer (US) > Offer (DS) ***
UC *vs.* UB	Promise (UC) > Promise (UB) *** Offer (UC) > Offer (UB) *	UC *vs.* DC	Promise (UC) > Promise (DC) *** Offer (UC) > Offer (DC) ***
UC *vs.* US	Promise (UC) > Promise (US) *** Offer (UC) $\approx$ Offer (US)	UB *vs.* DB	Promise (UB) > Promise (DB) *** Offer (UB) > Offer (DB) ***

† "$\approx$" denotes that the difference of the two averages is not statistically different from zero; ">" denotes that the difference of the two averages is statistically greater than zero. Standard errors are in parentheses. "*", "**" and "***" respectively denote significance at the 10%, 5% and 1% levels.

From the regression in Table 3, we find significant evidence for credibility, *i.e.*, a significant linear and positive relationship between promise and offer, independent of treatment conditions. This is confirmed by the coefficient on promise being significant at the one percent level. This credibility is also independent of the game

[10] Specifically, mean offers in UB is greater than that in DB (F-test, p-value = 0.0002), mean offers in US is greater than that in DS (F-test, p-value = 0.0000) and mean offers in UC is greater than mean offers in DC (F-test, p-value = 0.0000).

institution (dictator, ultimatum) and also the selection procedure (non-competitive, competitive). Thus, we can conclude that an aggregate offer–promise relationship exists that is independent of veto power and proposer selection and accordingly H2 is rejected. More specifically, a unit increase in promise, ceteris paribus results in a 0.27 unit increase in offer. This regression also shows that controlling for promise, the ultimatum conditions (UB, US and UC) have significantly higher offers than the dictator baseline (DB), though offers on the dictator treatments (DS, DC) do not significantly differ from the baseline (DB).

We further examine credibility (a positive relationship between promise and offer) by plotting offer against promise to determine whether any visual relationship can be observed. Scatter plots are given in Figure 1. From these, it does seem like there is a positive relationship, between offer and promise, which is more sharply defined for the ultimatum plots. This positive visual relationship is statistically confirmed by the positive and significant coefficient on promise in the regression on offers reported in Table 3.

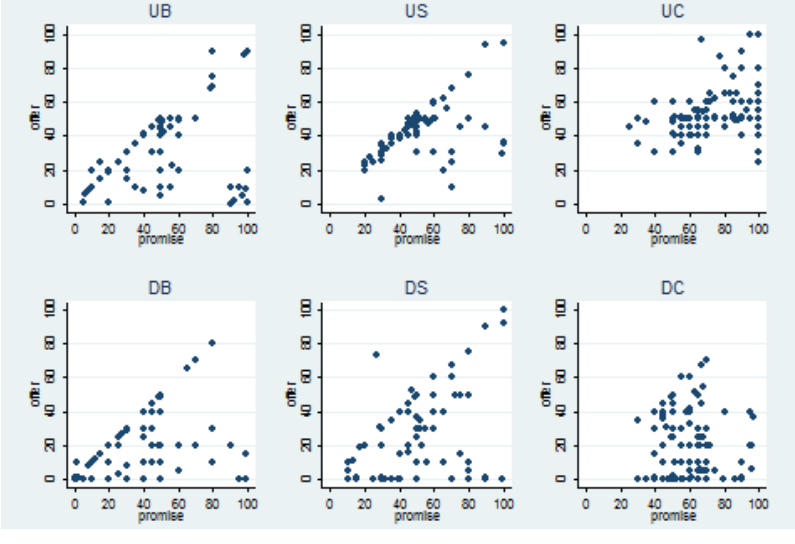


Figure 1. Scatterplots of offer on promise for all treatments.

In addition to identifying an offer–promise relationship in the aggregate, we also estimated offer promise relationships separately for all our conditions. In these regressions, not reported here, we find that offers are positively affected by promise at the five percent level for US and UC but no significant offer promise relationship for UB. Interestingly, the results from the dictator games are exactly in the opposite

direction and the DB condition displays credibility at the one percent level, whereas there is no significant offer promise relationship in DS and DC[11].

4.2. Credulity

We next explore the notion of credulity, *i.e.*,—recipients believe that a higher promise carries the intent of a higher offer and select the proposer who promises a higher amount.

Table 5 shows the averages of selected and refused promises in US, UC, DS and DC. The last columns 3 and 6 show the proportion of cases where the higher promise was selected, for UC and DC respectively[12]. Notice that, for our ultimatum treatments, the mean selected promise is greater than the mean refused promise, whereas forour dictator treatments, the selected promise is lower than the mean refused promise. To test whether this relationship holds statistically, we conducted an Ordinary Least Squares (OLS) regression for either condition of either game, of promise on a dummy variable which takes values 0 or 1, depending on whether the promise was refused or selected respectively. For the ultimatum game, the coefficient on this dummy showed that the selected promises were higher than the refused ones (p-value < 0.01 for both US and UC). For the dictator game, the reverse relationship was observed albeit with weaker statistical significance (p-value = 0.073 for DS, p-value = 0.05 for DC).

Table 5. Mean selected and refused promises.

	Ultimatum			**Dictator**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Selection	50.4 (100)	29 (2)	–	50.5 (101)	64.7 (7)	–
Competition	73.1 (120)	64.7 (120)	0.71 (111)	58 (125)	62.1 (125)	0.41 (119)

(1) mean selected promise (ultimatum); (2) mean refused promise (ultimatum); (3) proportion of high promise selection (ultimatum); (4) mean selected promise (dictator); (5) mean refused promise (dictator); (6) proportion of high promise selection (dictator); Number of observations are given in parentheses.

In either of our ultimatum treatment conditions, we can ask what the probability of any particular promise will be selected is, *i.e.*, estimate a selection function. This mapping gives the probability a promise will be selected as a function of the promise

[11] Why did credibility emerge in DB? One possibility is that the absolute power enjoyed by proposers in this condition paradoxically led to a manifestation of the preference for truth-telling, as in Vanberg [8], and thereby caused them to issue offers consonant with promises made, even if it did not actually lead to increased offers.

[12] Selectors in UC faced equal promises in 7.5 percent of cases while selectors in DC faced equal promises in 4.8 percent of cases.

itself. We investigate credulity by means of these functions, the aim also being to explore what the recipients infer about intent on receiving a promise.

For our estimation, we conducted a set of probit regressions where the value of the dependent variable was 0 or 1, depending on whether the promise was refused or selected, respectively. We suspected significant non-linearities in the data, and so used promise as well as promise exponentiated (up to the fourth degree) as independent variables. Likelihood ratio test results and the use of Akaike and Bayesian information criteria showed that the linear specification gives the best fit for UC. We found no estimable model for US, as very few promises were refused. The first column of Table 6 gives the constant and the coefficient on promise from the probit regressions using the linear specification for UC.

For the dictator games, the linear and cubic specifications respectively gave the best fits for DS and DC. The second and third column of Table 6 gives the constants and the coefficients on promise, promise squared and promise cubed, from the regressions using the appropriate specification for each condition. The selection function is not monotone increasing for DC. It is monotone decreasing for DS, with the caveat that the link between selection probability and promise is weak.

Table 6. Selection functions.

	UC (Linear)	**DS (Linear)**	**DC (Cubic)**
constant	−1.16 *** (0.42)	2.28 *** (0.75)	−7.55 *** (2.57)
promise	0.01 *** (0.005)	−0.02 * (0.009)	0.40 *** (0.14)
$Promise^2$		–	−0.006 *** (0.002)
$Promise^3$		–	0.00003 *** (0.00001)
Pseudo R^2	0.03	0.09	0.04
No. of obs.	240	90	250

Standard errors are in parentheses. "*", "**" and "***" respectively denote significance at the 10%, 5% and 1% levels.

The plot of selection probability (dependent variable) against promise (independent variable) can be derived using the estimated selection function above. Figure 2 depicts the selection function for UC which is monotonically increasing, *i.e.*, higher promises have greater likelihood of being selected. We could not, however, derive any such function for US. We thus find that for UC, hypothesis H3 is rejected, whereas, for DS, it cannot be rejected at the five percent level[13].

[13] In the DS regression, 18 observations corresponding to those from round 1 were dropped as they predicted selection perfectly.

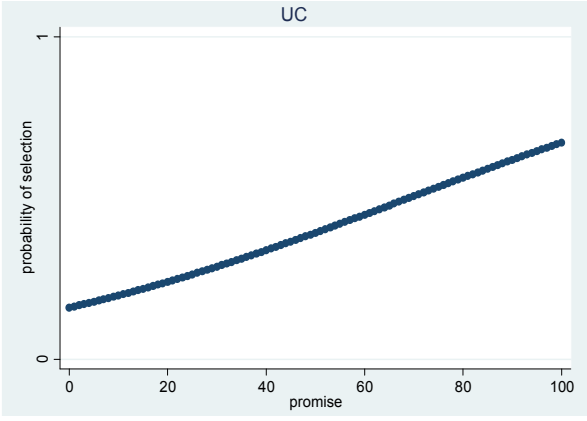


Figure 2. Selection function for UC.

4.3. Offer Rejection

Now, we turn to the issue of offer acceptance and rejection in the ultimatum game. Bearden [18] suggests a rejection rate of between 5 and 10 percent, while Oosterbeek *et al.* [19] arrive at a figure of 17 percent. While the literature analyzing rejection decisions is relatively scarce, available evidence suggests low offers, particularly those below 20 percent, are frequently rejected (see Camerer and Thaler [21]. We are specifically interested to see if H3 is rejected, *i.e.*,—recipients take into account not just offers but also prior promises, when making a decision to accept or reject an offer. Similar to earlier studies our data from UB, US and UC respectively show rejection rates of 13.5, 9 and 12.5 percent. We conducted probit regressions (not reported) on data pooled from condition pairs where the dependent variable is a dummy which takes values 0 or 1, depending on whether the offer was rejected or accepted respectively. The regressor was the relevant condition dummy. We found no difference in rejection rates across conditions[14]. The power on the part of the recipient to select her proposer hence induced no difference in the aggregate rate of rejection of offers.

No offer below 20 was recorded in UC. 23 percent of offers (22/96) were below 20 in UB. Of these, 40 percent (9/22) were rejected[15].By contrast, only five percent of the offers equaling or exceeding 20 (4/74) were rejected. For US, two percent of offers (2/100) were below 20. Both were rejected. On the other hand, only 7.1 percent

[14] The number of observations was 196, 216 and 220 for comparisons UB *vs.* US, UB *vs.* UC and US *vs.* UC respectively.

[15] One offer equaled 0 and was rejected.

of the offers equaling or exceeding 20 (7/98) were rejected. Results from UB and US thus mirror prior findings.

Average accepted offer in UB was 39.1, while average rejected offer was 13.5. The corresponding figures for US were 44.5 and 27.1 respectively. In UC, average accepted and rejected offers were respectively 53.7 and 37.1. For each condition, we conducted an OLS regression (not reported here) of offer on a dummy variable which takes values 0 or 1, depending on whether the offer was rejected or accepted respectively. The coefficient on the dummy was positive in each case (p-values less than one percent for every case). Hence, average accepted offer was greater than the average rejected offer, lending support to the contention that lower offers have a higher probability of rejection.

To determine what determines offer acceptance, we estimated acceptance functions which give the probability an offer will be accepted. We conducted probit regressions, one for each condition, where the dependent variable is a dummy which takes values 0 or 1, depending on whether the offer was rejected or accepted respectively. Apart from offer, the regressors were promise received for US, promise selected for UB, and promises selected and refused for UC. Results are given in Table 7[16,17].

We find, in line with usual findings, that offer is always used to determine acceptance, and that higher offers have lower chance of being rejected. Further, it is the only determinant in UB. Thus, prior promise is not used, in the absence of any selection, to decide rejection. Prior promise is used, however, in US, where offer and promise are both of importance in guiding the decision. Moreover, a higher promise, *ceteris paribus*, increases the probability of rejection.

Offer is the main determinant in UC. Both prior promises are additional determinants, with the likelihood of rejection rising in the prior selected promise, as in US, and falling in the prior refused one. A key finding is thus that recipients in US and UC reject an offer (and hence bear the cost of punishment) with greater likelihood if the prior attendant promise (on the basis of which the proposer was selected) is higher, while they do not in UB. Thus, H4 is rejected for environments with selection (US and UC) and offers that come from promissors that have a relatively high prior promise (independent of the offer made) face a higher probability of rejection.

---

[16] In the US, regression of 32 observations corresponding to those from rounds 1 and 4 were dropped as they predicted acceptance perfectly.

[17] We also ran probit regressions where we regressed offer on the gap between offer and promise (*i.e.*, offer–promise). The results from these regressions (not reported here) are in line with those in Table 7 confirming that the probability of acceptance of an offer is positively related to the gap between offer and promise, so, for example, if promise is higher than the eventual offer, then the gap is negative and the probability of acceptance becomes lower.

Table 7. Probit regression coefficients for the acceptance functions.

	UB	US	UC
offer	0.05 *** (0.01)	0.08 ** (0.03)	0.22 *** (0.05)
promise	−0.002 (0.004)	−0.02 ** (0.009)	−0.04 *** (0.01)
promise refused	–	–	0.03 ** (0.01)
constant	0.06 (0.67)	0.32 (0.94)	−7.57 *** (2.32)
Pseudo R^2	0.30	0.14	0.39
No. of obs.	96	68	120

Standard errors are in parentheses. "*", "**" and "***" respectively denote significance at the 10%, 5% and 1% levels.

5. Conclusions

The fact that the probability of the acceptance of an offer *ceteris paribus* decreases with an increase in the promised amount may be a primary driver of some of our other key results, namely, (i) why the selection function is monotone increasing in UC (credulity); (ii) why promises are credible in US and UC and not in UB; and (iii) why offers in US and UC are higher than in UB.

Our argument is as follows. The fear that an offer will be rejected if it is too low relative to promise made makes selected proposers choose offers consonant with promises, lending credibility to promises in US and UC. For UC, this assures recipients that higher promises are more likely to translate into higher offers, *i.e.*, makes them credulous, and thus renders higher promises more likely to be selected, and the selection function monotone. It is likely that a similar mechanism works in US as well, given results from Table 5; however, the paucity of refusal data does not permit us to statistically address the issue. In turn, this credulousness leads to promises being inflated due to selection pressure. Consequently offers are raised as well, due to credibility. In short, the use of prior promise as a factor in decision making at the offer rejection stage induces selected proposers to make offers close to the original promise, which causes recipients to select higher promises on average, which, in turn, generates selection pressure tending to inflate promises and thus offers.

Moreover, an offer is more likely to be accepted in UC if the promise the selector had refused earlier was higher. One plausible explanation for this rests on two premises: one, the use of the rejected promise as a statistic for the value of opportunity lost at the offer acceptance stage, and two, the evaluation of the value of the opportunity or offer in hand correlated with the one foregone. That is, promise refused earlier is used at the offer acceptance stage by the recipient as a signal of the offer that would have accrued if that alternative had been selected. Hence, the refused promise provides the value of foregone opportunity, and perhaps a reflection of regret (see, e.g., Bell [22] and Loomes and Sugden [23]. At the same time, the recipient correlates her evaluation of the offer in hand with her evaluation of the opportunity lost. Hence, a higher (lower) value of promise refused increases

(reduces) the value of the opportunity foregone to her, and, thereby, raises (lowers) the value of the offer in hand, thus increasing (reducing) the likelihood of acceptance.

As far as the comparison of offers between US and UC is concerned, it is possible that competition is the main driving factor. The presence of a competing promise causes each potential proposer to raise his promise, and this in turn raises offers in UC, relative to US, due to credibility of promises. Additionally, the fact that refused promise is used in UC to guide the rejection decision may contribute to offers being higher in UC because it provides the recipient with an extra instrument which implicitly raises her bargaining power vis-à-vis that of the proposer. No such relationship is obtained, however, in the baseline condition UB where, in the absence of selection, promises hold little meaning. This finding is consistent with their being a conditional preference for hearing truth, as expressed through rejection behavior, the condition being that any such preference is triggered only when promise was used earlier as the basis for selection.

In conclusion, the power to select her partner on the basis of promise can generate benefits to the recipient in the ultimatum game, tilting the division of surplus in her favor, especially with competitive selection. This is because recipients reject an offer with greater likelihood when the prior promise associated with it is higher, *ceteris paribus*. This behavior, which could be a reflection of some form of a preference for hearing truth, endogenously induces credibility with respect to promise and credulity with respect to offer. The absence of rejection power in the dictator game severs the link between punishment and promise, and thereby does not impact offers, and fails to produce credibility or credulity.

Our findings may contribute to the larger inquiry into how non-binding communication can convey intent, or how communication can be credible in social dilemmas, dating back at least to Loomis [24]. Casual empiricism suggests partner choice usually exists, recourse is often accessible and partnerships are regularly entered into on the basis of promises, sometimes with successful consummation. It seems reasonable to then suppose that communication is delivering some intent. In such contexts, our findings point to the possibility that outcome may be linked to communication's capacity to convey intent, and, moreover, such capacity may be endogenously activated if recipients use its content as a factor when deciding on the posterior recourse, given it has been used for the prior selection decision. In other words, the use of communication when deciding on prior partner selection endows it with hysteresis, and renders it relevant for posterior decisions as well, thereby enabling it to convey intent.

This essay's explorations were confined to ultimatum and dictator environments, and to selection based on communication. Examining connections between promises and partnership formation and performance in other incentive domains such as trust games, prisoners' dilemma and public good games, or coordination games,

and selection through alternate means such as contests are left for future research. Another possible avenue for further study is whether the communicational hysteresis identified above extends to wider settings.

Acknowledgments: We are grateful to seminar participants at University of Calcutta, Jawaharlal Nehru University, Indian Statistical Institute, National Institute of Public Finance and Policy, Indian Institute of Management, Bangalore, and Indian School of Business for helpful comments. We are indebted to Sarmila Banerjee, Ananish Chaudhuri, Birendra Rai, Shubhro Sarkar, Christoph Vanberg, Robert Veszteg, and an anonymous referee for their questions and suggestions.

Author Contributions: All of the authors contributed equally to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A: Instructions

Instruction for DB

Today, you will play several rounds of a game.In each round, the entire set of students will be divided into many groups, each group consisting of one *proposer* and one *recipient*. Each group plays within itself during a round, and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are two stages: the *promise* stage, and the *offer* stage.

The proposer makes a promise in the first stage.

In the second stage, the proposer has a cake with him of size 100. He decides how much to give to the recipient (integer between and including 0 and 100). That amount is the recipient's payoff for that round. The amount the proposer keeps back (100—amount given) is his payoff for that round.

In the first stage, the proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will give in the second stage. However, promise is not binding: the proposer is free to give more or less than his promise.

Total payoff at the end is the sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Instruction for UB

Today, you will play several rounds of a game. In each round, the entire set of students will be divided into many groups, each group consisting of one *proposer* and

one *recipient*. Each group plays within itself during a round, and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are three stages: the *promise* stage, the *offer* stage, and the *decision* stage.

The proposer makes a promise in the first stage.

In the second stage, the proposer has a cake with him of size 100. He decides how much to offer to the recipient (integer between and including 0 and 100).

The recipient can reject in the third stage. If she rejects, both parties get 0 payoff. If she accepts in the third stage, the amount offered is the recipient's payoff for that round. The amount the proposer keeps back (100—amount offered) is his payoff for that round.

In the first stage, the proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will offer in the second stage. But promise is not binding: the proposer is free to offer more or less than his promise.

Total payoff at the end is sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Instruction for DS

Today you will play several rounds of a game. In each round, the entire set of students will be divided into many groups, each group consisting of one *proposer* and one *recipient*. Each group plays within itself during a round and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are three stages: the *promise* stage, the *selection* stage, and the *offer* stage.

The proposer makes a promise in the first stage.

The recipient decides whether to select or refuse in the second stage. The game goes to the third stage if there is selection in the second stage.

In the third stage, the proposer has a cake with him of size 100. He decides how much to give to the recipient (integer between and including 0 and 100). That amount is the recipient's payoff for that round. The amount the proposer keeps back (100—amount given) is his payoff for that round.

The game may not go to the third stage—if the recipient refuses in the second stage. In that case, both parties get 0 payoff. Selection/Refusal is on the basis of a promise.

In the first stage, the proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will give in the third stage if selected. However, promise is not binding: once selected, the proposer is free to give more or less than his promise.

Total payoff at the end is sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Instruction for US

Today, you will play several rounds of a game.In each round the entire set of students will be divided into many groups, each group consisting of one *proposer* and one *recipient*. Each group plays within itself during a round, and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are four stages: the *promise* stage, the *selection* stage, the *offer* stage, and the *decision* stage.

The proposer makes a promise in the first stage.

The recipient decides whether to select or refuse in the second stage. The game goes to the third stage if there is selection in the second stage.

In the third stage, the proposer has a cake with him of size 100. He decides how much to offer to the recipient (integer between and including 0 and 100).

The recipient can reject in the fourth stage. If she rejects, both parties get 0 payoff. If she accepts in the fourth stage, the amount offered is the recipient's payoff for that round. The amount the proposer keeps back (100—amount offered) is his payoff for that round.

The game may not go to the third stage—if the recipient refuses in the second stage. In that case, both parties get 0 payoff. Selection/Refusal is on the basis of a promise.

In the first stage, the proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will offer in the third stage if selected. However, promise is not binding: once selected, the proposer is free to offer more or less than his promise.

Total payoff at the end is sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Instruction for DC

Today, you will play several rounds of a game. In each round the entire set of students will be divided into many groups, each group consisting of two *proposers* and one *recipient*. Each group plays within itself during a round, and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are three stages: the *promise* stage, the *selection* stage, and the *offer* stage.

The proposers each make a promise in the first stage.

The recipient decides which one to select in the second stage. The other is refused.

In the third stage, the selected proposer has a cake with him of size 100. He decides how much to give to the recipient (integer between and including 0 and 100). That amount is the recipient's payoff for that round. The amount the proposer keeps back (100—amount given) is his payoff for that round.

Selection/Refusal in the second stage is on the basis of a promise.

In the first stage, each proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will give in the third stage if selected. But promise is not binding: once selected, the proposer is free to give more or less than his promise.

Total payoff at the end is sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Instruction for UC

Today, you will play several rounds of a game. In each round, the entire set of students will be divided into many groups, each group consisting of two *proposers* and one *recipient*. Each group plays within itself during a round, and has no interaction with any other group.

The selection of groups is done randomly by the experimenter. At the end of any round, all groups are broken up and new groups are formed for the next round by randomly rematching the participants. The interaction in any round is anonymous.

There are four stages: the *promise* stage, the *selection* stage, the *offer* stage, and the *decision* stage.

The proposers each make a promise in the first stage.

The recipient decides which one to select in the second stage. The other is refused.

In the third stage, the selected proposer has a cake with him of size 100. He decides how much to offer to the recipient (integer between and including 0 and 100).

The recipient can reject in the fourth stage. If she rejects, both parties get 0 payoff. If she accepts in the fourth stage, the amount offered is the recipient's payoff for that round. The amount the proposer keeps back (100—amount offered) is his payoff for that round.

Selection/Refusal in the second stage is on the basis of a promise.

In the first stage, each proposer sends a promise (integer between and including 0 and 100), interpreted as how much the proposer will offer in the third stage if selected. But promise is not binding: once selected, the proposer is free to offer more or less than his promise.

Total payoff at the end is sum of payoffs over rounds. Cash compensation will be higher if total payoff is higher. You can earn between 100 and 500.

Appendix B: Response Sheets

Response Sheet for DB

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____
Stage II:
Proposer's actual offer: _____

Response Sheet for UB

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____
Stage II:
Proposer's actual offer: _____
Stage III:
Recipient's decision: Accept/Reject

Response Sheet for DS

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____

Stage II:
Recipient's decision: Select/Refuse
Stage III:
Proposer's actual offer: _____

Response Sheet for US

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____
Stage II:
Recipient's decision:Select/Refuse
Stage III:
Proposer's actual offer: _____
Stage IV:
Recipient's decision: Accept/Reject

Response Sheet for DC

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____
Stage II:
Recipient's Subject No.: _____
Response:Select/Refuse
Stage III:
Proposer's actual offer: _____

Response Sheet for UC

Round #:
Stage I:
Proposer's Subject No.: _____
Promised amount: _____
Stage II:
Recipient's Subject No.: _____
Response:Select/Refuse
Stage III:
Proposer's actual offer: _____
Stage IV:
Recipient's decision:Accept/Reject

References

1. Hoffman, E.; McCabe, K.; Shachat, K.; Smith, V. Preferences, Property Right and Anonymity in Bargaining Games. *Games Econ. Behav.* **1994**, *7*, 346–380.
2. Navarro, N.; Veszteg, R. Demonstration of Power: Experimental Results on Bilateral Bargaining. *J. Econ. Psychol.* **2011**, *32*, 762–772.
3. Gale, J.; Binmore, K.; Samuelson, L. Learning to be Imperfect: The Ultimatum Game. *Games Econ. Behav.* **1995**, *8*, 56–90.
4. Roth, A.; Prasnikar, V.; Okuno-Fujiwara, M.; Zamir, S. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study. *Am. Econ. Rev.* **1991**, *81*, 1068–1095.
5. Cox, C. Inequity Aversion and Advantage Seeking with Asymmetric Competition. *J. Econ. Behav. Organ.* **2013**, *86*, 121–136.
6. Bochet, O.; Putterman, L. Not Just Babble: A Voluntary Contribution Experiment with Iterative Numerical Messages. *Eur. Econ. Rev.* **2009**, *53*, 309–326.
7. Ellingsen, T.; Johannesson, M. Promises, Threats and Fairness. *Econ. J.* **2004**, *114*, 397–420.
8. Vanberg, C. Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica* **2008**, *76*, 1467–1480.
9. Tullock, G. Non-Prisoner's Dilemma. *J. Econ. Behav. Organ.* **1999**, *39*, 455–458.
10. Coricelli, G.; Fehr, D.; Fellner, G. Partner Selection in Public Goods Experiments. *J. Confl. Resolut.* **2004**, *48*, 356–378.
11. Brandts, J.; Cooper, D.; Weber, R. Legitimacy, Communication and Leadership in the Turnaround Game. *Manag. Sci.* **2015**, *61*, 2627–2645.
12. Dulleck, U.; Kerschbamer, R.; Sutter, M. The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition. *Am. Econ. Rev.* **2011**, *101*, 526–555.
13. Beck, A.; Kerschbamer, R.; Qiu, J.; Sutter, M. Shaping Beliefs in Experimental Markets for Expert Services: Guilt Aversion and the Impact of Promises and Money-Burning Options. *Games Econ. Behav.* **2013**, *81*, 145–164.
14. Goeree, J.; Zhang, J. Communication and Competition. *Exp. Econ.* **2014**, *17*, 421–438.
15. Forsythe, R.; Horowitz, J.; Savin, N.; Sefton, M. Fairness in Simple Bargaining Experiments. *Games Econ. Behav.* **1994**, *6*, 347–369.
16. Güth, W.; Schmittberger, R.; Schwarze, B. An Experimental Analysis of Ultimatum Bargaining. *J. Econ. Behav. Organ.* **1982**, *3*, 367–388.
17. Heston, A.; Summers, R.; Aten, B. *Penn World Table Version 7.0.*; Center for International Comparisons of Production, Income and Prices, University of Pennsylvania: Philadelphia, PA, USA, 2011.
18. Bearden, J. Ultimatum Bargaining Experiments: The State of the Art. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=626183 (accessed on 14 January 2016).
19. Oosterbeek, H.; Sloof, R.; van de Kuilen, G. Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Exp. Econ.* **2004**, *7*, 171–188.
20. Engel, C. Dictator Games: A Meta Study. *Exp. Econ.* **2011**, *14*, 538–610.

21. Camerer, C.; Thaler, R. Anomalies: Ultimatums, Dictators and Manners. *J. Econ. Perspect.* **1995**, *9*, 209–219.
22. Bell, D. Regret in Decision Making under Uncertainty. *Oper. Res.* **1982**, *30*, 961–981.
23. Loomes, G.; Sugden, R. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Econ. J.* **1982**, *92*, 805–824.
24. Loomis, J. Communication, the Development of Trust, and Cooperative Behavior. *Hum. Relat.* **1959**, *12*, 305–315.

MDPI AG
Klybeckstrasse 64
4057 Basel, Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
http://www.mdpi.com/

Games Editorial Office
E-mail: games@mdpi.com
http://www.mdpi.com/journal/games

MDPI AG
Klybeckstrasse 64
4057 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

Basel • Beijing • Wuhan • Barcelona

ISBN 978-3-03842-206-8