Chai, Yidan (2022) *Information on social value orientation and cooperation in public goods games.* Doctor of Philosophy (PhD) thesis, University of Kent,.

The version of record is available from
https://doi.org/10.22024/UniKent/01.02.93228

This document version
UNSPECIFIED

DOI for this version

INFORMATION ON SOCIAL VALUE ORIENTATION AND COOPERATION IN PUBLIC GOODS GAMES

DISSERTATION

Submitted in Partial Fulfillment of

the Requirements for

the Degree of

DOCTOR OF PHILOSOPHY (Economics)

at the

UNIVERSITY OF KENT
SCHOOL OF ECONOMICS

by

Yidan Chai

September 2021

INFORMATION ON SOCIAL VALUE ORIENTATION AND COOPERATION IN PUBLIC GOODS GAMES

DISSERTATION

Submitted in Partial Fulfillment of

the Requirements for

the Degree of

DOCTOR OF PHILOSOPHY (Economics)

at the

UNIVERSITY OF KENT
SCHOOL OF ECONOMICS

by

Yidan Chai

September 2021

Approved:

____

Department Chair Signature

____

Date

University ID: 17904360

Approved by the Guidance Committee:

Major: Economics

Advisor's Name

Date

Advisor's Name

Date

Advisor's Name

Date

Advisor's Name

Date

Acknowledgements

First and foremost, I would like to thank school of Economics and Wuhan University for the financial and technique support for the first experiment. I am extremely grateful to Prof. Edward Cartwright for his financial support for my second and third experiments.

I would like to thank my esteemed supervisors Prof. Edward Cartwright, Dr. Fernanda Leite Lopez de Leon, and Dr. Alexander Klein, for their essential counsel, unwavering support, and patience during my PhD program. Their vast expertise and wealth of experience have aided me throughout my academic career and daily life. Especially, Prof. Edward Cartwright's invaluable support in shaping my experiment methods and critiquing my results is something I would like to express my gratitude for.

I would also like to thank Dr. Lian Xue for her support and advice on my study. My gratitude extends to the Faculty of school of Economics. It is because of their kind assistance and support that my studies and life in the UK have been so enjoyable. Finally, I would like to thank my parents, my cousin, and friends for their support. It would have been difficult for me to finish my studies without their wonderful understanding and encouragement throughout the last few years.

Yidan Chai

September 2021

To my family

ABSTRACT

INFORMATION ON SOCIAL VALUE ORIENTATION AND COOPERATION IN PUBLIC GOODS GAMES

by

Yidan Chai

Advisor: Prof. Edward Cartwright, Ph.D.

Supervisors: Dr. Fernanda Leite Lopez de Leon, Ph.D.

Dr. Alexander Klein, Ph.D.

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Economics)

September 2021

In the present thesis, we report the design and results of three experiments. These experiments focused on the impact of information of social value orientation (SVO) in public goods game on cooperative behavior in public goods game.

To provide a theoretical underpinning to the work we apply belief-based guilt aversion and reciprocity models from psychological game theory. These allow us to derive testable hypotheses on the role of information on contributions.

The abstracts of the experiments are as below.

Experiment 1: We explore the extent to which information about the social value orientation (SVO) of group members influences contributions to a good project. We compare four networks of information - empty, pair, star and complete. We find that information about SVO has a significant impact on contributions in an experimental setting. This impact is consistent with theoretical predictions based on reciprocity and guilt aversion. We show that a star network with a pro-social manager has the biggest positive effect on contributions.

Experiment 2: We explore whether information on own social value orientation (SVO) impacts on contributions to a public good. We argue that if pro-social behavior is driven by belief-based preferences then private information on SVO should not impact on contributions; but if behavior is driven by internalised norms then information on own SVO should impact contributions. We report an experiment with three treatments: no information on SVO, binary information whether pro-self or pro-social, and SVO indicated on a scale from very pro-social to very pro-self. We observe no effect of information on contributions. This finding supports a belief based approach of modelling pro-social behavior.

Experiment 3: We explore the influence of leader's social value orientation on follower and group contributions to a public good. We offer the opportunity for leaders to signal their SVO types by messages before followers making contribution decisions. Three treatments are conducted in terms of the ways of message sending: Cheap-talk, Hide and Truth. In the Cheap-talk game, leaders can choose to tell followers they are pro-social, pro-self or to hide their types (deception opportunity available); in the Hide game, leaders can only choose to reveal their true SVO type or to hide their types (no deception); in the Truth game, instead, computer makes the revealing decision – send true type message or hide the type (no deception, no right to choose). We find that the majority of pro-social leaders choose to send a message revealing their true type, while the majority of pro-self leaders prefer to reveal themselves as pro-social or just hide their SVO type. In terms of the effect of messages, we find that contributions are higher if the leader sends a pro-social message, than if the message does not reveal type or is pro-self. Moreover, we find that followers react to the message rather than the credibility of the message. Finally, for the group overall efficiency, we suggest that the optimal game or mechanism to use depends on the likely type of the leader.

The structure of the present thesis is arranged as following. Chapter 1 introduces theoretical basis of the experiments, including social value orientation, public goods game, psychological game theory and reciprocity. Chapters 2, 3 and 4 elaborate the three experiments conducted.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Sociality is the characteristic that organisms show when they act as individuals in collective activities or as members of society that are conducive to collective and social development. It is inevitable for an individual to interact with other people in social activities. For instance, employees execute their duty and realize diverse functions in their firm, where individual job function is necessarily linked with others. Productivity is, thus, higher is employees are willing to cooperate with each other on group tasks.

As a matter of fact, as discussed in Section 1.1 below, the population is heterogeneous with regard to their preferences or values towards social interactions. To one extreme, some people are willing to benefit collective interest regardless of whether other people cooperate or not, while to another extreme some people only care about their own interest and even pursue larger relative gain by decreasing others' benefits.

It is thus of interest to consider the consequences for cooperation of heterogeneity within the group. And also to question how people react to information about the cooperative tendency of others in the group. In the thesis we ran three experiments designed to explore how cooperation in a group is influenced by information about own and others pro-sociality:

- In Experiment 1, reported in Chapter 2, we varied the information that a person has on the pro-sociality of others. We find that information has a systematic and significant effect on cooperation and efficiency. Information increases or decreases cooperation depending on the composition of the group, i.e. number of pro-social or pro-self members of the group.

- In Experiment 2, reported in Chapter 3, we looked at the impact of information on own pro-sociality. We find that information has a small but significant effect on contributions.

- In Experiment 3, reported in Chapter 4, we looked at the incentives and consequences of people being able to hide or lie about own pro-sociality. We find that pro-self people are likely to lie or hide type but this ultimately increases efficiency because it creates a frame of pro-sociality.

Before outlining the experiments in detail the present chapter provides some preliminaries. We measured pro-sociality using social value orientation (SVO) and so Section 1.1 reviews the literature around (SVO). The setting used to model cooperation is a linear public goods game and so Section 1.2 reviews the relevant literature. Finally, we use a theoretical framework based around psychological game theory to derive testable hypotheses for each of the experiments. This framework is introduced in Section 1.3.

1.1 Social Value Orientation

The theory for Social Value Orientation (henceforth SVO) has been recognized and gradually developing since the 1960s in psychology [Messick and McClintock, 1968]. The notion is utilized broadly in many research areas as well as in practice, such as industrial employment activities. The SVO classification is designed to capture diverse inclinations and evaluations towards self-interest and others' interest when individuals are making their decisions in an interactive environment. Specifically, to capture personal preferences in decision-making processes where actions, not only affect own interests, but also others.

There are mixed dispositional concerns among the population as to the notion of fairness and social preferences [Van Lange, 1999]. This suggests that the primary consideration of own payoffs is not the only factor affecting actions, and that actions would also be influenced by personality and the consequences of others. Theoretically, these individual dispositional concerns are measured by distinct weights assigned towards one's own and other people's material payoff [De Cremer and Van Lange, 2001]. These weights are typically measured by an SVO angle.

According to the SVO angle, some incipient literatures have divided SVO into diverse types, such as generosity, martyrdom, masochism and aggression [e.g. McClintock and van Avermaet, 1982]. Nowadays, three key types have emerged within the SVO framework, and are commonly employed in recent researches. The first classification is **Pro-social**, a type of people who regard as important others' welfare as well as themselves. Thus they care about the equality and joint welfare before taking actions. Alternatively, a person is **Pro-self** which can be split into two types. One is Individualist, which indicates that this type of people only concern about their self-interest but pay no attention on welfare of other people. We also have Competitors who are pursuing larger relative benefits in interdependent situations. In this case, they would better-off when the difference between their own and others' welfare is intensified. Therefore, they would always seek an action which would cause greater advantageous disparity compared to others' benefits [Smeesters et al., 2003].

An open question is the extent to which SVO is a fundamental and stable personal preference [Messick and McClintock, 1968]. Some have argued that SVO might change with age, and additionally as people learn from repeated interactions with others, or on account of other external factors [Van Lange et al., 1997b]. Evidence also shows that SVO was easily affected by framing [De Dreu and McCusker, 1997]. Smeesters et al. [2002] found that individuals with low consistent orientations would behaved more unstable over time than people who have high consistency. Moreover, the influence of SVO is stronger on people with consistent SVO, and weaker on those with less consistent SVO who are affected more by situational information instead.

1.1.1 SVO on Behaviour

Experimental evidence indicates that SVO does has an influence on individual's cooperative behaviour along with the expectations about others' actions [McClintock and Liebrand, 1988]. Generally speaking, pro-self individuals behave more collaboratively than pro-selfs in social dilemmas [Kramer et al., 1986, McClintock and Liebrand, 1988, Van Lange and Liebrand, 1989, De Cremer and Van Lange, 2001, Simpson, 2004]. Studies have also found that pro-socials have an inclination to enhance the equality of the group and behave in a conditionally cooperative way [Van Lange, 1999].

How can the connection between SVO and cooperative behaviour in social dilemmas be explained? Bogaert et al. [2008] proposed a conceptual model of individual decision-making process. In their model, people's SVO has an initial influence on their basic willingness to cooperate, and as well an impact on their rudimentary expectations of partner's behaviour with regard to reciprocity. Then, a specific interdependent situation which contains contextual cues will moderate their original willingness and expectations into context-specific ones. Eventually the moderated cooperative goal and expectations will mediate their final behavioural decisions. In all, this conceptual model shows how basal willingness and expectations can be moderated and then in turn mediate their actual cooperative behaviour given a specific situation.

In terms of the expectations of others, studies suggest that pro-selfs expect less cooperation than pro-socials [Kelley and Stahelski, 1970, Liebrand, 1984, Van Lange and Liebrand, 1991, Van Lange, 1992, Smeesters et al., 2003]. In otherwords individuals have a tendency to think others are 'like them'. It should be highlighted, however, that the differences between pro-socials and pro-selfs in both behaviour and beliefs, while statistically significant, are not always large in economic terms. Moreover, external cues can moderate the effect of SVO. For example, pro-selfs can be more motivated to cooperate if there is a strong sense of group identity [De Cremer and van Dijk, 2002]. Or they can see strategic benefits [van Dijk et al., 2004]. Besides, pro-socials may be more sensitive to external information about others and, in a reciprocal way, lower contributions [De Cremer and Van Lange, 2001].

1.1.2 Measurement of SVO

Several methods have been used to assess individuals' SVO types. A full review can be found in Murphy and Ackermann [2014]. A primary method is to elicit the types of individuals based on their actual behaviour observed. If an individual acted cooperatively, they would be classified as a pro-social. However, as we have discussed in the previous section about the conceptual model, natural preferences are easily moderated by environmental information, thus the observed behaviour usually cannot fully demonstrate individuals' true SVO types. In general, this method lacks the ability to distinguish between internal and moderated inclinations of subjects. Thus, the SVOs classified with this method are found to have lower explanatory power in different interactive situations.

The most prevalent method is an assessment using the decomposed games technique. Among this technique, there are various structures of assessments. The first one is the Triple-Dominance Measure (TDM) [Van Lange et al., 1997b]. In this assessment process, subjects need to select a most preferred distributions of own-other payoffs from three options in each of nine questions. Every option indicates an allocation of points between themselves and another person. At the end of the assessment, people will be classified into one of three SVO types: pro-social, individualistic or competitive. A related structure of assessment is called as the Ring Measure [Liebrand, 1984]. In this measurement, the preferences of each subject is represented by a vector starting from the origin of a rectangular coordinate system, with own payoff as horizontal axis and other's payoff as vertical axis. Points in the space demonstrate four different SVO types.

The TDM and ring measure have been widely used by economists. However, they have limitations as described by Murphy et al. [2011]. In particular, they have low-resolution and lack sensitivity to personal diversities. This possibly stems from the methods being somewhat non-intuitive to complete. Murphy et al. [2011] proposed a new SVO Slider Measure. This measure gives researchers more difference-sensitive outputs with higher explanation power, which can be used into potential comparisons and as a dependent variable to investigate more meticulous correlation between influence factors and personal preferences. The method involves making allocation choices sequentially in six simulated dictator games.[1] After the six decisions are made an SVO angle can be calculated and individuals classified into SVO type. In our experiments we use the slider measure.

1.1.3 Why SVO in This Thesis

One might ask, why we employ information on SVO as a factor in our experiments, rather than other psychological trait or some demographic trait like gender, ethnicity. It is important for us to explain it thoroughly. First, the idea can suit. The concept of SVO is used to describe and categorise people according to their personal attitudes about the distribution of resources. The concept broadly distinguishes between two categories of decision makers: People with a more team-oriented approach are referred to as "pro-social" and those with a more egocentric focus are described as "pro-self". The idea can suit our setting of public goods games, where resources are distributed within teams. Moreover, there are contradicted best outcomes (a dilemma) for teams and for individuals, faced by the players. Second, the question that we like to address in our experiments is how the information on SVO of other players impact individuals' decision makings.

[1] An additional nine items can be used as a secondary test for separation of joint outcome maximization and inequity aversion

SVO is a clear notion that players can understand quickly and establish beliefs upon it – Pro-social is cooperative and pro-self is the group who are selfish. In this sense, gender or ethnicity do not have a very clear image for players to build beliefs in terms of cooperation preferences. Besides, the study does not focus on the gender effect or ethnicity effect, we are more focus on player's beliefs caused by the expectations on different psycho-social types. Last but not least, why not other psychological concepts? Research revealed that social value orientation exhibited considerable ability to predict actual behaviours in a variety of different experiment games, such as donation, cooperative behaviour in social dilemmas, with pro-social exhibiting greater cooperation than individualists and competitors, as what discussed in above sections. Therefore, we think SVO is one of the suitable constructs that we could use in our experiments.

1.2 Public Goods Game

The structure of a typical linear PGG can be viewed as a simplified activity where members of a group can voluntarily contribute towards a public good. The non-excludable and non-rivalrous nature of the public good means that all group members can consume or utilize the public good regardless of the amount they contribute.

More specifically, assume that there are n players in a set N and they will interact with each other as a group for one round of a PGG. At the beginning of the round, each of them is given E tokens of private good as an endowment. Players are then required to determine how many tokens they would contribute into the collective account of the group. Denote the contribution from each participant i as

a_i, where $i \in N$ and $a_i \in [0, E]$. Once all players have contributed, the tokens in the group account are accumulated and multiplied by a positive multiplier $\alpha > 1$. Then, each member of the group will receive a return from the public good equal to α/n. The payoff for each group member is shown as following:

$$u_i(a_1, \ldots, a_n) = E - a_i + \frac{\alpha}{n} \sum_{i=1}^{n} a_i \tag{1.1}$$

The expression $\frac{\alpha}{n}$ indicates the marginal per-capita return from the public good. We can also denote it as $m \in (0, 1)$. The expression $\sum a_i$ is the total amount of tokens contributed to the public good, we can also denote it as A.

The individual incentive of a player in this one shot PGG is to contribute nothing to the public good. This is because the return from contributing to the public good is less than the return from keeping the private good $m < 1$. Therefore, in the Nash equilibrium there would be zero amount of tokens in the group account and each subject earns nothing but their endowment E. However, total payoffs in the group are maximized if all players contribute their full endowment. In this case each player obtains payoff $\alpha E > E$. A PGG, thus, involves a social dilemma of individual interests v.s. group interests. Of interests to economists is how to ameliorate free-riding and stimulate a higher level of contribution in PGG.

The following paragraphs briefly summarize some pertinent experimental evidence from PGG studies. For a comprehensive review of the literature see Chaudhuri [2011]. Firstly, there are heterogeneous levels of contribution among the population, ranging from free-riding to almost complete cooperation (e.g. 83% in Isaac et al. [1984]). Within which, most participants indeed contributed something rather than purely free-riding as predicted by individual incentives. Marwell and Ames [1980] found that there was a weaker effect of free-riding than theoretical predictions in one-shot PGG, where average contribution rates in their works were both exhibited at around half of the endowment. Similarly, the rate was at around 50% of cooperation in the first round of repeated PGG too, where participants would interact with each other over several rounds [Mark Isaac et al., 1985, Isaac et al., 1984, Kim and Walker, 1984].

Secondly, a majority of participants are conditional cooperators, which means that their cooperative behaviour is conditional on actions taken by others. As we have discussed in the last section, with regard to SVO, the actions taken and motives thereof, and the underlying intentions before actions both matter in terms of the cooperative behaviour. Moreover, other-regarding individuals are concerned about both the distribution of payoffs and fairness. Experimental evidence justifying the evidence for conditional cooperation in PGG includes Fehr and Schmidt [1999], Keser and van Winden [2000], Fischbacher et al. [2001], Gächter [2006], Fischbacher and Gächter [2010].

Thirdly, though there is a high initial rate of contributions, in a repeated PGG the average level of contributions declines over time and typically attains a low level of cooperation in the last round [e.g. Mark Isaac et al., 1985, Andreoni, 1988, Fehr and Schmidt, 1999]. This decline in cooperation could be a result of conditional cooperation and subjects lowering their contribution as they observe the low contributions of others. Various mechanisms have been shown to increase contributions and reduce the decline in contributions over time. For instance, when either non-monetary or monetary punitive opportunity is available, cooperation cooperation tends to be high and maintains a high level [Fehr and Schmidt, 1999, Fehr and Gächter, 2000b, Masclet et al., 2003].

Fourth, the framing of a PGG can influence contributions [Gächter et al., 2017].

The literature has looked, in particular, at the contrast between give and take, and public good versus public bad. Framing can influence behaviour in PGG by changing beliefs and, therefore, influencing conditional cooperation [Dufwenberg et al., 2011, Ellingsen et al., 2012].

My thesis contributes to this literature in various respects. Specifically, information about the SVO of members in the group may directly influence contributions by providing information to make a more informed decision, and it may indirectly influence contributions through changing the frame of reference. For instance, if player i is labelled as pro-social then both i and others who get to know the type of i may contribute differently to if i is labelled pro-self. We, thus, see how heterogeneity of types, coupled with changes of frame, can influence contributions. The present thesis discusses this in more detail in individual chapters. The theoretical framework uses is based on psychological game theory.

1.3 Psychological Game Theory

The basic idea behind psychological game theory is that a player's payoff in a game may depend on their beliefs and not just the actions of others Geanakoplos et al. [1989]. In a public good game a player may feel guilty if they contribute less than they believe others expected them to. Similarly, they may want to reciprocate what they believe others will do. Psychological game theory is relevant for our work (explained in detail in Chapters 2-4) because it is a way of modelling how information on SVO may influence beliefs which then influence contributions. The following paragraphs briefly introduce guilt aversion and reciprocity.

The first detailed model of guilt in extensive games is proposed by Battigalli and Dufwenberg [2007]. The theory builds on the intuitive premise that when people behave in a way that goes against the positive expectations of others, they will feel guilty. The degree of guilt is related to the difference between the payoff others expected to receive and the payoff they will receive with one's own behaviour. Guilt will reduce one's own utility. The basis of this framework can be traced back to the comment made by Baumeister et al. [1994].

To briefly introduce the general model of guilt aversion consider a game with player set N a game tree with set T of nodes. Let s_i denote a strategy of player i and s_{-i} the strategy of other players. We then have a set H_i of all information that player i has on all nodes. Element h_i of H_i is a specific information set for player i. Player i can condition their strategy s_i on the information h_i they receive in the game.

Based on the received and updated information h_i, player i develops his/her *First Order Belief* $b_i(j|h_i)$. This represents the player's belief about the strategies of other players $j \in N$. For instance, in a public good game the contributions of others. Player i also develops his/her *Second Order Belief* $c_i(j|h_i)$ which is a belief about $b_j(i|h_j)$ for other $j \in N$. In other words, second order belief is about self and what a person believes others think about him/her.

The expected monetary payoff $E(u_j)$ of player j in the game can be formed on basis of his/her First Order Belief. We can then calculate the payoff gap between expectation and reality as the 'disappointment' of player j: $D_j(s_i, s_{-i}) = \max\{0, E(u_j) - u_j(s_i, s_{-i})\}$, where $u_j(s_i, s_{-i})$ denotes realized payoff. Let $G_{ij} = D_j(z) - \min_s D_j(z(s, s_{-i}))$ be player i's contribution to j's disappointment.

Player i's final expected payoff is then:

$$U_i(s_i, s_{-i}, b_i, c_i) = m_i(s_i, s_{-i}) - \sum \theta_{ij} G_{ij}(\cdot) \tag{1.2}$$

Where θ indicates the level of guilt aversion. Thus, player i feels guilt if he/she lets down the payoff expectations of others. This guilt depends on second order beliefs.

There are several noteworthy applications of belief-based guilt aversion theory in psychological games. For instance Charness and Dufwenberg [2006] provide evidence of guilt aversion. They found that communication and promises influence both participants' First Order and Second Order Beliefs in simple principal-agent games. This means that promises are positively correlated with cooperation. Reuben et al. [2009] also find laboratory evidence supporting the important role that guilt aversion plays in trust games. A systematic introduction of the theory and related studies can be found in the survey by Cartwright [2019].

The evidence, however, is support of belief-based guilt aversion is not universally positive. For example, Vanberg [2008] emphasized the experiment results discussed above in Charness and Dufwenberg [2006] could be explained differently. In particular, the experiment design fails to separate personal preferences for keeping promises from guilt at disappointing another. In lights of these flaws, Vanberg added a swap treatment where participants would be matched with a recipient who had received a promise from someone else. General behaviour did not significantly differ across treatments. These findings suggest the existence of a commitment-based account for dictator's behaviour regardless of others' expectations; players have preferences not to lie instead of feeling guilty when disappointing people.

In other studies, guilt aversion theory has been tested further. Ellingsen et al.

[2010] found no evidence for guilt aversion in their dictator game. However, there was some evidence which was suggestive of guilt aversion for some belief choice sets in the experiment. Kawagoe and Narita [2014] proposed a new version of guilt aversion which is called as personal guilt aversion, suggesting that a feeling of guilt emerges when subjects abandon their counterpart's expectations thus their promised. The experiments yield no strong correlation between actions and elicited beliefs, which goes against the predictions of belief based guilt aversion. Charness and Dufwenberg [2010] themselves also find limited support for belief based guilt aversion.

The literature connecting belief-based guilt aversion and public goods game will be discussed in detail in the chapters to follow. At this stage we highlight two closely related papers results that are supportive of the belief-based approach. Dhami et al. [2019] propose a model with reciprocity, guilt-aversion and surprise-seeking and find that guilt aversion is pervasive. Caria and Fafchamps [2017] conducted public goods game on a star network, where the treatment group has an expectation-revealing setting. The experimental evidence support the idea that such disclosure has a positive effect on contributions from central players on the network, consistent with guilt aversion, although only in groups with weak social ties outside the experiment. The weak social ties is argued in the study saying that participants in this situation would underestimate expectation others have toward them.

1.4 Reciprocity

In this section, a brief introduction to the concept of reciprocity in psychological game theory is presented. Rabin [1993] incorporates a concept of fairness into economic models by extending the framework from Geanakoplos et al. [1989] to a two player game. In the model, three elements are included – player i's own action $a_i \in S_i$, his belief about other's action b_i and a second order belief about other's belief about his action c_i. The 'kindness' level of player i to player j then can be written as following:

$$f_i(a_i, b_j) \equiv \frac{\pi_j(b_j, a_i) - \pi_j^e(b_j)}{\pi_j^h(b_j) - \pi_j^{min}(b_j)} \tag{1.3}$$

Where π^e denotes the average between the highest and lowest payoff player j could have (among Pareto-efficient points) given b_j. Similarly, π^h denotes the highest payoff, and π^{min} means the lowest possible payoff. Player i is, thus, kind (or unkind) if he/she believes her actions result in player j obtaining a payoff above (or below) the average possible.

Then, player i's belief on the kindness of player j is written as following:

$$\widetilde{f}_j\,(b_j, c_i) \equiv \frac{\pi_i(c_i, b_j) - \pi_i^e(c_j)}{\pi_i^h(c_i) - \pi_i^{min}(c_i)} \tag{1.4}$$

The expected utility function of player i consists of his/her kindness and perceived kindness level, shown as below:

$$U_i(a_i, b_j, c_i) \equiv \pi_i(a_i, b_j) + \widetilde{f}_j\,(b_j, c_i) \times [1 + f_i(a_i, b_j)] \tag{1.5}$$

Thus, player i would want to be kind if he/she believes player j was kind and would want to be unkind if he/she believes player j was unkind.

Dufwenberg and Kirchsteiger [2004] structure a model concentrating on reciprocity in sequential games. They also slightly revise the notion of kindness and allow for many players. Now we have player i's own action $a_i \in A_i$, his belief about j's action $b_ij \in B_{ij}$ and a second order belief about j's belief about his action

$c_i ji$.The kindness level of player i is described as the difference to average of highest and lowest material payoff of player j

$$\kappa_{ij} = \pi_j(a_i, b_{ij}) - \frac{1}{2} \times [\max\{\pi_j | a_i \in A_i\} - \min\{\pi_j | a_i \in E_i\}] \tag{1.6}$$

where E_i is the set of efficient actions. Further, player i's beliefs about player j's kindness to themselves is defined as:

$$\lambda iji = \pi_i(b_{ij}, c_{ijk}) - \frac{1}{2} \times [\max\{\pi_j | b_{ij} \in B_{ij}\} - \min\{\pi_j | b_{ij} \in E_{ij}\}] \tag{1.7}$$

Hence, the utility of player i can be written as following, where Y_{ij} – sensitivity to reciprocity – is an exogenous positive value given and the second part of the function representing as the reciprocity payoff:

$$U_i = \pi_i(a_i, b_{ij}) + \sum (Y_{ij} \cdot \kappa_{ij} \cdot \lambda_{iji}) \tag{1.8}$$

The reciprocity model thus consists of two parts: the final monetary payoff that player i obtains, and a 'payoff' measured by perceived kindness level of intended actions of others. The cooperation thus is conditional on how perceived kindness of others are. For an alternative formulation see Falk and Fischbacher [2006].

There is a large literature applying psychological game theoretical models of reciprocity and tested its predictions. For example, in the paper by Bhirombhakdi and Potipiti [2011], they test the prediction performance of the model from Dufwenberg and Kirchsteiger [2004]. They found that the model has a good performance comparing to other models, where the prediction accuracy is overall 80% in different experimental scenarios. Falk et al. [2008] test the relevance of fairness intentions on observed behaviour. Based on their experimental data, they suggest that models – focusing on both intentions and monetary distributions – can well-capture the behaviour. Thus, in their opinion, the pure belief-based models proposed by Rabin [1993] and Dufwenberg and Kirchsteiger [2004] are not complete in terms of the factors included.

In experiments conducted by Dhaene and Bouckaert [2010], based on the model by Dufwenberg and Kirchsteiger [2004], they classify participants into two patterns of behaviour according to their first and second order beliefs. The findings from data collected in sequential prisoner's game are broadly in line with the predictions of belief-based reciprocity, as well as those collected in previous studies. However, it turns out that around 80% of the first movers are acting 'too kindly' which are beyond the predictions of belief-based reciprocity.

Another recent study proposes a further factor to consider in reciprocity based models of intentions and perceived motives. Orhun [2018] considers reciprocal intention is driven by two elements: the expectations of final outcomes and voluntariness. Though, experimental evidence support the idea that a higher voluntariness would lead to higher positive reciprocity, they found existing models fail to adequately account for kindness perceptions. Their data shows that if the first mover in the game is perceived to be reacting to strategic incentives then positive reciprocity is lower.

A recent study by Isoni and Sugden [2019] also distinguish reciprocity into two parts – a reciprocal kindness which is utilized in the model by Rabin [1993], as well as a reciprocal cooperation which capture trust in interactive games. They argue that the model by Rabin [1993] fails to appropriately model trust leading to a 'Paradox of trust'. The model by Dufwenberg and Kirchsteiger [2004] eliminates the problem but is still unable to explain behaviour in trust games.

Overall, therefore, we see that a complex mix of factors need be accounted for to model behaviour. Even so, there is strong evidence that belief-based motives – including guilt aversion and reciprocity – are an important part of the mix. Information about SVO can be expected to change beliefs and, thus, influence behaviour. This is why we focus on belief-based models in the thesis. In doing so we draw, in particular, on the work of Dufwenberg et al. [2011] who apply reciprocity and guilt aversion to public goods game. They demonstrate that changes in framing can influence beliefs and behaviour. Their work is discussed in more detail in Chapter 2.

1.5 Overview of Subsequent Chapters

In next chapters, three experiments are presented. Chapter 2 shows public goods games with four different public information networks; Chapter 3 continues the spirit of the design which investigates the effects of private information on provision in public goods games; Chapter 4 focuses on more detailed aspect which researches on the influence of leader's information on the contribution.

Specifically, the information of players' social value orientation types were provided within groups in public goods games. This action is the main intervention in all experiments that we conducted. Chapter 2 has four information networks: Empty, Complete, Star and Pair – four ways of public information providing. We compared the complete network, where every player knows every other players SVO, the empty network, where a player only knows own SVO, a pair network, where each player knows about one other group member's SVO, and a star network, where leader's SVO is known by other group members. Chapter 3 controls the private information disclosure – whether contributions influenced if players know their own SVO. Chapter 4 provides such information of leaders to followers in groups. We gave leaders the chance to reveal, lie about or hide their SVO from followers. This allowed us to study whether there is 'strategic revelation' of information.

We believe that various ways of providing information make the framing in the game different, which will further affect the player's belief and behavior. This hypothesis is based on our model and previous theoretical research and experimental results [Dufwenberg et al., 2011, Ellingsen et al., 2012, Gächter et al., 2017]. At the same time, based on previous research, SVO is in a stable state in the short term [Messick and McClintock, 1968]. If the player's behavior changes, it can be explained that it is sensitive to external information. On the one hand, it can show that our intervention is effective. On the other hand, it can also prove that the player's behavior is more affected by the external environment than following the principle of personal quality [De Cremer and Van Lange, 2001, De Dreu and McCusker, 1997].

Generally, information that a player is pro-social will make players to have a tendency to increase contributions, and information that a player is pro-self to decrease contributions. The overall effect will depend on the balance of pro-socials and pro-selfs in the group and other factors, such as the type of information revealed and demographics. Overall, we find good support for hypotheses based on guilt aversion and reciprocity. We also, however, find some evidence that is hard to reconcile with a belief-based models.

In the first experiment, we found that there is a natural difference between pro-socials and pro-selfs where the former are more cooperative. In terms of the impact of providing public information, we found that, after being informed the types of other group members, pro-socials reacted more cooperatively, but pro-selfs acted in an opposite way. Moreover, we found that both pro-socials and pro-selfs' cooperation was increasing with the number of pro-socials in their group, especially in the complete treatment. Also, the effect of information disclosure on group contributions differs with the structures of network. Contributions are highest in the complete network. Contributions in the star network are highly dependent on whether the 'star' is pro-social or pro-self.

In the second experiment, we again found that pro-social contribute significantly more than pro-selfs. Here the effect was more pronounced than in the first experiment. We also found that there is a small effect of SVO information revelation on a player's own behaviour. But, it shows that pro-selfs are more sensitive to information of their type and then to be less cooperative. When we use the classification obtained in the second stage to study the behavior of the first stage, we found that the Unconditional Cooperators and Free-riders are not stable in terms of cooperation when exposing with more external information. Binary information impacts these types negatively most. Moreover, the type of leader has greater effect on group total contribution than the effect resulting from the types of followers in the group, where a pro-self leader induces lower level of contribution in the group account. In terms of the type from test results and self-valued types, we found that pro-socials who have pro-social belief of themselves are more significantly cooperative comparing to other types. Furthermore, a pro-social information stimulates the contribution of self-valued pro-selfs.

Finally, in the third experiment in Chapter 4, when it comes to sending messages, we found that the majority of pro-social leaders would like to deliver a message that reflects their real personality trait. The majority of pro-self figures, on the other hand, tend to conceal their true type by revealing themselves as pro-social or simply hiding their SVO type. Also, we look at the impact of the type of leader. The contribution of followers is shown to be higher (for a given leader contribution) when the leader type is pro-social than when the leader type is unknown or whether the leader is pro-self. Furthermore, we found that pro-selfs view unknown leaders as if they were pro-self, while pro-socials have more favorable reactions to unknown leaders. Besides, we found that when the leaders appeared to be pro-social, followers did not participate more on an individual basis in the Truth setting than in the Cheap-talk setting (for a fixed leader contribution). Last, we run a group simulation for each session and found that the total group contributions are smaller in the Cheap-talk environment than in the Truth setting.

Our experimental results further support the findings that pro-socials are natural cooperators compared to pro-selfs [Kramer et al., 1986, McClintock and Liebrand, 1988, Van Lange and Liebrand, 1989, De Cremer and Van Lange, 2001, Simpson, 2004]. Especially, results in Chapter 2 show that the public information of others or leaders can affect individual interactive behaviour to an extent [Gächter et al., 2017], and they also indicate that anticipation of guilt can motivate players to be more cooperative [Chang et al., 2011]. Therefore, we experimentally see that a reasonable intervention of revealing SVO (or other psychological personality traits) would encourage people interact in a more cooperative way. Interestingly, our findings show that players have expectations of aftermath of their actions – majority of pro-self leaders in the third experiment choose to lie or hide their true SVO type. The finding riches the researches regarding to beliefs and expectations in interactive games [Ackermann and Murphy, 2019, Murphy and Ackermann, 2013, Gächter and

Renner, 2010].

Chapter 2

Information on social value orientation and contributions to a group project: Theory and experimental evidence

[0]Work joint with Edward Cartwright (De Montfort University, UK), Lian Xue (Wuhan University, China)

[0]Each researchers' contribution: the idea, original experimental design and coding, experiment conducting, major parts of analysis by Yidan; theory, model suggested and help by Prof. Cartwright; Dr Xue helped experiment conducting, data analysis as well.

[0]Funding for this project was provided by the Center for Behavioral and Experimental Research, Wuhan University, China. & PhD Research Allocation, School of Economics, University of Kent, UK.

2.1 Introduction

It is relatively routine for employees in large organisations to undergo some form of psychometric testing, either at interview or while on the job [Sackett and Walmsley, 2014]. One basic rationale is to manage the psychological composition of working groups in order to increase productivity and cooperation [Belbin, 2011, 2012].[1] The act of taking a psychometric test and knowing the results are available to others may, though, influence behaviour. For instance, someone who is 'labelled' as non-cooperative may behave less cooperatively in a group project as a consequence. Similarly, someone who knows that another member of their group is non-cooperative may cooperate less because of an expectation of lower cooperation in the group [Drouvelis and Georgantzis, 2019]. In this experiment we analyse whether information about the social value orientation of group members influences behaviour in a public goods game.

The notion of social value orientation (SVO) has developed over many years within psychology [Messick and McClintock, 1968, McClintock and van Avermaet, 1982, Van Lange, 1999]. The basic idea is to measure an individuals willingness to share money with (or take money from) another. In common usage we can distinguish pro-socials, who are willing to forego own payoff to benefit others, and pro-self, who are either only concerned about own payoff (individualist) or who are willing to forego own payoff to get ahead of others (competitors). A large number of studies have explored the relationship between SVO and contributions in public good games [e.g. Perugini et al., 2005, Fiedler et al., 2013, Hoyer et al., 2014, Böhm and Theelen, 2016, Mill and Theelen, 2019]. As one might expect there is

[1] Whether or not this actually works is an open question (Partington and Harris [1999], Edwards et al. [2007]).

evidence that pro-socials contribute more to the public good than pro-selfs [Balliet et al., 2009, Pletzer et al., 2018]. It is interesting to note, however, that the effect is often small (and sometimes non-existant) [Renkewitz et al., 2011]. Moreover, with repeated interaction social ties, built through shared experience, come to the fore [Van Dijk et al., 2002, Sonnemans et al., 2006]. SVO, and in particular the classification into pro-social and pro-self, is, therefore, a relatively noisy indicator of expected contributions.

Despite SVO being a noisy indicator of contributions, we hypothesised that information about the SVO of others would have a significant and predictable effect on contributions. To model the influence of SVO information we use models of guilt aversion and reciprocity taken from psychological game theory [Dufwenberg and Kirchsteiger, 2004, Battigalli and Dufwenberg, 2007, 2009, Dufwenberg and Patel, 2017, Battigalli and Dufwenberg, 2019, Cartwright, 2019]. Specifically, we follow the approach of Dufwenberg et al. [2011] in positing that an individuals contribution in a public good game will depend on her first order beliefs over others' contributions (because of reciprocity) and second order beliefs over what she believes others expect her to contribute (because of guilt aversion). We then assume that knowing the SVO of others may influence her first order beliefs, while knowing that others know her SVO may influence her second order beliefs. We have, therefore, precise channels through which information about SVO may influence contributions.

To study the effect of information we analyze different networks of information within the group. We say that there is an undirected link between two group members if both are informed of each other's SVO. We compare different network structures including the Complete network, where there is a link between any pair of members in the group, the Star network, where there exists a 'manager' that is linked to everyone else in the group, and a Pair network where group members are each linked with one other member. Studying network structure is interesting from an applied point of view because it can, in principal, be exogenously controlled by the authority that forms the group. It also proves interesting at a theoretical level. Specifically, it seems natural that more information an individual has about the SVO of others the more it influences her first order beliefs. Similarly the more other members of the group know her SVO the more it influences her second order beliefs. This allows us to derive testable hypotheses relating network structure to public good contributions.

The effect of information will fundamentally depend on what that information is. For instance, an individual informed that another group member is pro-self may decrease her first order beliefs on the contributions of others. Conversely, if she is informed the group member is pro-social she may increase her first order beliefs. Given that an approximate 50-50 split between pro-selfs and pro-socials is typical one might expect that on average information cancels out. We find, however, that information has a net positive impact on contributions. The primary reason for this is that pro-socials contribute significantly more when others can observe their SVO (and increase their contribution more than pro-selfs decrease theirs). According to our theoretical model this points to the importance of guilt aversion for pro-socials. Our results support, therefore, the notion that anticipation of guilt can motivate cooperative behaviour [Chang et al., 2011].

Our results give interesting insight on how organisations could use the information from psychometric tests to increase cooperation. At a basic level we reinforce the idea that psychometric tests are probably not a great measure of how an individual will perform in a team. In our setting, pro-selfs contributed as much as pro-socials. If, however, information about SVO is widely disseminated across the group this can increase contributions to the group project. We find that the optimal network structure is the Star network with a pro-social manager. In this case contributions increase from a baseline of around 50% of the maximum (with no information) to 70% or more. The Star network is advantageous because you only need one pro-social member of the group to make a big difference, provided that member is made the manager. We discuss the implications of this more in the Conclusion.

In relating our work to the prior literature we being by noting that several studies have looked at public good games on a network [e.g. Cassar, 2007, Bramoullé and Kranton, 2007, Carpenter et al., 2012, Falk et al., 2013, Boosey, 2017]. Another strand of literature has looked at how behaviour is influence by information each individual has on the contributions of others [e.g. Andreoni and Petrie, 2004, Bigoni and Suetens, 2012, Hartig et al., 2015]. Let us highlight, therefore, that in our work we keep the underlying public good network and information on individual contributions constant throughout. This allows us to isolate the effect that information about SVO has on behaviour. Various studies show that 'information' about other group members can influence behaviour. For instance, Andreoni and Petrie [2004] and Samek and Sheremeta [2014] find that photos of subjects increases contributions to the public good. Similarly, information that allows identification with the group (e.g. all from the same university) has been shown to increase contributions, especially for those identified as pro-social [De Cremer and Van Vugt, 1999, De Cremer et al., 2008].

The only study we are aware of that systematically varies information on personality traits is Drouvelis and Georgantzis [2019]. They compare, in two player interaction, the consequences of subjects being informed, and the person they are matched with, are above or below average in terms of agreeableness. Like us they find that information influences cooperation. Although in their case the net effect is negative with information on disagreeableness pulling down contributions. Our work complements their study in two key respects. First, we provide a theoretical framework, based around psychological game theory, to analyse the effect of information about personality. Second, use of a four player public good game allows us to systematically explore changes in the network of information.

We proceed as follows: In Section 2.2 we introduce the model and notation. In Section 2.3 we provide a theoretical analysis based on guilt aversion and reciprocity. In Section 2.4 we explain our experiment design and in Section 2.5 provide the results. Section 2.6 concludes with additional discussion of the related literature. Additional materials including the full experiment instructions are contained in an Appendix.

2.2 Model and Notation

Consider a standard linear public good game. There is a set of players $N = \{1, ..., n\}$. Each player is endowed with E units of private good. The players simultaneously and independently decide how much of their endowment to contribute to a public good. Let $a_i \in [0, E]$ denote the contribution of player $i \in N$. Let $a = (a_1, ..., a_n)$ denote the contribution profile and $A = \sum_{i=1}^{n} a_i$ denote total contributions. The monetary payoff of player i is given by

$$u_i = E - a_i + mA \tag{2.1}$$

where $m \in (1/n, 1)$ is the marginal per-capita return from the public good. Given that $m < 1$ a player motivated solely by own monetary payoff has no incentive to contribute to the public good. Given, however, that $m > 1/n$ it is Pareto efficient for all players to contribute their full endowment to the public good.

2.2.1 Guilt aversion and reciprocity

Following the approach of Dufwenberg et al. [2011] we consider the role of guilt aversion [Charness and Dufwenberg, 2006, Battigalli and Dufwenberg, 2007] and and reciprocity [Rabin, 1993, Dufwenberg and Kirchsteiger, 2004]. Let b_{ij} be player i's first order belief about j's contribution. Specifically, it is the mean of the probability distribution that player i has over the contribution of player j. Let c_{iji} denote player i's second order belief about b_{ji}. Specifically, it is the mean of the probability distribution that player i has over the first order belief of player j. Let $B_i = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} b_{ij}$ denote player i's average first order belief on the contributions of others. Let $C_i = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} c_{iji}$ denote player i's average second order belief.

With a model of simple guilt, player i feels guilt if she contributes less to the public good than she believes other players expected her to [Battigalli and Dufwenberg, 2007]. So, her payoff can be written

$$u_i^g = E - a_i + mA - \gamma_i max\{0, C_i - a_i\} \tag{2.2}$$

where $\gamma_i \geq 0$ measures the extent of player i's guilt aversion. Note that if $m + \gamma_i > 1$ it is optimal for player i to contribute C_i. Thus, guilt aversion can incentivize a player to contribute to the public good. Indeed, if she is sufficiently guilt averse ($\gamma_i > 1 - m$) and believes others expert her to contribute her full endowment

$(C_i = E)$ then it is optimal for her to contribute E.

With a model of reciprocity player i wants to reciprocate the kindness, or unkindness of others. Following Dufwenberg et al. [2011] the payoff of player i can be written

$$u_i^r = E - a_i + mA + Y_i \left(a_i - \frac{E}{2}\right)\left(B_i - \frac{E}{2}\right) \tag{2.3}$$

where $Y_i \geq 0$ measures the strength of player i's desire to reciprocate. Note that if $B_i > \frac{E}{2}$ and Y_i is sufficiently large then it is optimal for player i to contribute E. Thus, reciprocity can also incentivise a player to contribute to the public good. The intuition for this result is that if others are kind by contributing then player i reciprocates this kindness by also contributing.

In the following we will consider a general formulation

$$u_i^*(a_i, A, B_i, C_i) = E - a_i + mA - \gamma_i max\{0, C_i - a_i\} + Y_i \left(a_i - \frac{E}{2}\right)\left(B_i - \frac{E}{2}\right)$$

This formulation allows that player i could experience both guilt aversion and reciprocity. This approach seems natural given the evidence that both guilt aversion and reciprocity may drive behaviour in public good games [Dufwenberg et al., 2011]. This could be because some people experience both guilt aversion and reciprocity and/or because some people experience guilt aversion and others reciprocity.

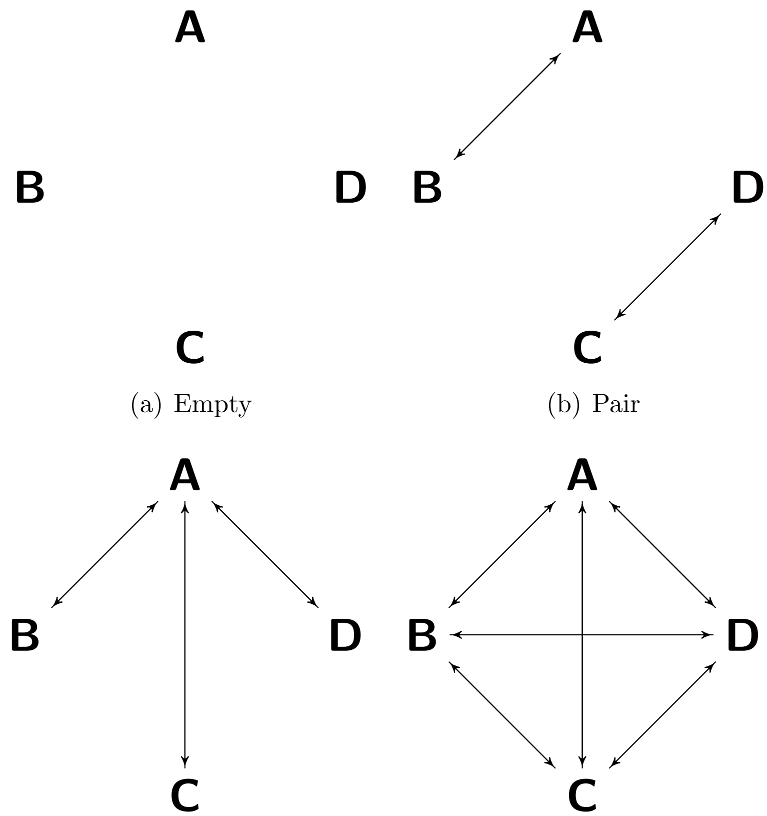
2.2.2 Network of information

We assume that each player has taken a test to measure their social value orientation (SVO). On the basis of this test each player is categorized as either pro-social or pro-self. The details of the test and categorization are not crucial at this point. Let v_i denote the SVO of player i where $v_i = 0$ denotes pro-self and

$v_i = 1$ denotes pro-social. For now we make no direct connection between a player's SVO, v_i, and her guilt aversion, γ_i, and strength of reciprocity, Y_i.

There exists a network of information about SVO within the set of players. This is represented by adjacency matrix G. In interpretation if $G_{ij} = 1$ then player i is told the SVO of player j, while if $G_{ij} = 0$ player i does not know the SVO of player j. We assume throughout that the network is non-directed meaning that $G_{ij} = G_{ji}$ for all $i, j \in N$. We also set $G_{ii} = 1$ for all $i \in N$ meaning that player i knows her own SVO. Let $G_i \subset N/\{i\}$ denote the set of players with whom player i is linked (other than herself). Let $g_i = |G_i|$ denote the number of players about whom player i knows the SVO. Note that g_i also denotes the number of players who know the SVO of player i. We let f_i denote the number of players in set G_i who are pro-self and $l_i = g_i - f_i$ denote the number who are pro-social.

We refer to the network of information as *empty* if $G_{ij} = 0$ for all $i, j \in N, i \neq j$. We refer to the network as *complete* if $G_{ij} = 1$ for all $i, j \in N$. There are two other 'intermediate' networks we will pay particular attention to. We refer to a *pair* network if $g_i = 1$ for all $i \in N$. In other words, each player has information on one other player and so the group is composed of pairs of players who have information about each other. Finally, we refer to a *star* or *manager* network if there exists player $i \in N$ such that $g_i = n - 1$ and $g_j = 1$ for all $j \in N, j \neq i$. In interpretation, we think of player i as the manager who has information on every player. Others, merely have information about the manager. These four different networks are illustrated in Figure 2.1 for a group of size 4 (with players A, B, C and D).



(a) Empty (b) Pair

(c) Star (d) Complete

Figure 2.1: Four examples of a network of information

2.3 Theoretical Results

Each player $i \in N$ can be characterised by list $Z_i = (v_i, Y_i, \gamma_i, B_i, C_i)$ detailing whether player i is categorized as pro-social, her reciprocity and guilt sensitivity and her first and second order beliefs. The following result is a straightforward characterisation of the optimal strategy. Crucially, we see that the optimal strategy depends on first and second order beliefs.

Proposition 2.1. *If player i maximizes her payoff then she will contribute: (i) E if $m + Y_i(B_i - E/2) > 1$, (ii) C_i if $1 > m + Y_i(B_i - E/2) > 1 - \gamma_i$, (iii) 0 otherwise.*

Proof. We consider each possibility in turn. (i) If $m+Y_i(B_i-E/2) > 1$ then we must have $B_i > E/2$. So, $du_i^*/da_i \geq -1+m+Y_i(B_i-E/2) > 0$. Hence it is optimal to contribute the maximum E. (ii) If $a_i > C_i$ then $du_i^*/da_i = -1+m+Y_i(B_i-E/2) < 0$. So, it cannot be optimal to contribute more than C_i. If $a_i < C_i$ then $du_i^*/da_i = -1+m+\gamma_i+Y_i(B_i-E/2) > 0$. So, it cannot be optimal to contribute less than C_i. (iii) The only remaining possibility is $m+Y_i(B_i-E/2)+\gamma_i < 1$. If $a_i < C_i$ then $du_i^*/da_i = -1+m+\gamma_i+Y_i(B_i-E/2) < 0$. So, it is optimal to contribute 0. □

We conjecture that information about others will influence first and second order beliefs and, thus, influence contributions through guilt aversion and reciprocity. Before we explore this in detail we introduce a preliminary assumption. This assumption is stronger that we need to derive our results but provides an intuitive benchmark. It encapsulates two ideas: (1) Symmetry: if player i has the same information about players j and k then she has the same beliefs about j and k. (2) Independence: Player i's beliefs about player j are not influenced by what she knows about some other player k (or what j may know about player k).

Assumption 2.1. *(a) For any player $i \in N$ there exists real numbers b_i^0, b_i^l, b_i^f such that for any $j \in N, j \neq i$ (i) if $g_{ij} = 0$ then $b_{ij} = b_i^0$, (ii) if $g_{ij} = 1$ and $v_j = 1$ then $b_{ij} = b_i^l$, and (iii) if $g_{ij} = 1$ and $v_j = 0$ then $b_{ij} = b_i^f$. (b) For any player $i \in N$ there exists real numbers c_i^0, c_i^1 such that for any $j \in N, j \neq i$ (i) if $g_{ij} = 0$ then $c_{iji} = c_i^0$, and (ii) if $g_{ij} = 1$ then $c_{iji} = c_i^1$.*

An assumption of symmetry is natural in a neutral, anonymous setting, such as the experimental laboratory, where a player has no additional information except SVO to inform beliefs. Within an organization additional information may exist to inform beliefs, e.g. past experience working with someone. In a large organisation, however, anonymity may still be a reasonable approximation. The assumption of independence is slightly stronger. For instance, one could imagine that a player would update her beliefs about the population distribution of SVO based on observing a sample of SVOs. Independence, therefore, requires a player to have a well informed prior on the population distribution of SVO (and expect others to have a well informed prior). This requirement could be relaxed without any substantive change in our results, provided that observing the SVO of some does not lead to a radical shift in beliefs about the population distribution.

2.3.1 Guilt Aversion

We can now explore the effect that information can have on beliefs and, therefore, contributions. Our basic conjecture is that players who are pro-social are expected to contribute more and those who are pro-self are expected to contribute less. We begin the formal analysis by looking at second-order beliefs.

Assumption 2.2. *For any player $i \in N$, (a) if player i is pro-social then $c_i^1 > c_i^0$, and (b) if player i is pro-self then $c_i^1 < c_i^0$.*

Assumption 2.2 implies that if a player is pro-social then her second-order belief is increasing in the number of players with whom she is linked. Conversely if she is pro-self her second-order belief is decreasing in the number of players with whom she is linked. This leads directly to our first main result. For simplicity we state this in terms of guilt aversion by setting $Y_i = 0$. We consider the general case shortly.

Proposition 2.2. *(a) If $Y_i = 0$ and player i is pro-social then her contribution is increasing in g_i if $\gamma_i > 1 - m$ and independent of g_i otherwise. (b) If $Y_i = 0$ and player i is pro-self then her contribution is decreasing in g_i if $\gamma_i > 1 - m$ and independent of g_i otherwise.*

Proof. Consider network of information G. Note that g_i denotes the number of players who know i's SVO and $n - 1 - g_i$ denotes the number who do not. Given Assumption 2.1 the average second-order belief of player i is given by

$$C_i = \frac{1}{n-1}\left(g_i c_i^1 + (n-1-g_i)c_i^0\right).$$

Applying Assumption 2.2 we see that C_i is increasing in g_i if i is pro-social and decreasing in g_i if i is pro-self.

Now consider Proposition 2.1 and a ceteris paribus change in C_i. Second-order beliefs are only relevant in scenario (ii) and here we have $a_i = C_i$. Hence an increase in C_i leads to an increase in a_i if $1 > m + Y_i(B_i - E/2) > 1 - \gamma_i$ and no change otherwise. Similarly a decrease in C_i leads to a decrease in a_i if $1 > m + Y_i(B_i - E/2) > 1 - \gamma_i$ and no change otherwise.In this case we can set $Y_i = 0$ to derive the conditions in the statement of the Proposition. □

The intuition behind Proposition 2.2 is relatively straightforward. If a player is pro-social then she expects that others would expect her to contribute relatively a lot. So, if she is pro-social, and guilt averse, the more people know she is pro-social then the more she contributes in order to avoid guilt. Similarly, if a player is pro-self then she expects that others would expect her to contribute relatively little. So, if she is pro-self, and guilt averse, the more people know she is pro-self then the less she can contribute and avoid guilt.

2.3.2 Reciprocity

We turn our attention now to first-order beliefs. Here we assume that a player expects those who are pro-social to contribute relatively more and those who are pro-self to contribute relatively less.

Assumption 2.3. *For any player $i \in N$ we have $b_i^l < b_i^0 < b_i^f$.*

Assumption 2.3 implies that a player's first-order belief is increasing in the number of players with whom she is linked who are pro-social, l_i, and is decreasing in the number of players with whom she is linked who are pro-self, f_i. This leads directly to our second main result. For simplicity we state this in terms of reciprocity by setting $\gamma_i = 0$.

Proposition 2.3. *(a) If $\gamma_i = 0$ then player i's contribution is weakly increasing in l_i. (b) If $\gamma_i = 0$ then player i's contribution is weakly decreasing in f_i.*

Proof. Consider network of information G. Note that g_i denotes the number of players who know i's SVO and $n-1-g_i$ denotes the number who do not. Let f_i denote the number of players in set G_i who are pro-self and $l_i = g_i - f_i$ denote the number who are pro-social. Given Assumption 2.1 the average first-order belief of player i is given by

$$B_i = \frac{1}{n-1}\left(f_i b_i^f + (g_i - f_i) b_i^l + (n-1-g_i) b_i^0\right).$$

Applying Assumption 2.3 we see that B_i is increasing in l_i and decreasing in f_i.

Now consider Proposition 2.1 and a ceteris paribus change in B_i. If

$$B_i > \frac{1-m}{Y_i} + \frac{E}{2}$$

then $a_i = E$. If

$$\frac{1-m}{Y_i} + \frac{E}{2} > B_i > \frac{1-\gamma_i - m}{Y_i} + \frac{E}{2}$$

then $a_i = C_i$. Otherwise $a_i = 0$. This is sufficient to prove the Proposition and also motivate the examples to follow in the text. □

To illustrate Proposition 2.3 we will provide one example. (see the proof of Proposition 2.3 for the details.) We introduce term

$$\bar{b} = \frac{1-m}{Y_i} + \frac{E}{2}.$$

Suppose that $b_i^0 < \bar{b} < b_i^l$. In this case, if G is the Empty network then player i will contribute 0, because $b_i^0 < \bar{b}$. By contrast, if G is the Complete network and everyone (possibly excluding i) is pro-social then player i will contribute E, because $\bar{b} < b_i^l$. More generally, if player i observes enough players in the group who are pro-social then she will switch from contributing 0 to contributing E. She does so because the information forms her to revise upwards her first-order belief and, because of positive reciprocity, contribute herself.

2.3.3 General setting

We move now to the general case in which we allow for both guilt-aversion and reciprocity. To do so we introduce some terminology. Consider player $i \in N$. When we refer to a *ceteris paribus increase in* f_i this can be interpreted as either keeping l_i constant (so g_i increases) or keeping g_i constant (so l_i decreases). Similarly where we refer to a *ceteris parbius increase in* l_i this can be interpreted as either keeping f_i or g_i constant. The following result follows immediately from Propositions 2.2

and 2.3.

Proposition 2.4. *(a) If player $i \in N$ is pro-social then her contribution is, ceteris paribus, weakly increasing in l_i. (b) If player i is pro-self then her contribution is, ceteris paribus, weakly decreasing in f_i.*

Proof. We know from the proof of Proposition 2.2 that if player i is pro-social her contribution is weakly increasing in g_i. We know from Proposition 2.3 that her contribution is weakly increasing in f_i. Part (a) of the Proposition follows immediately. A similar logic treats part (b). □

Proposition 2.4 treats the case where guilt aversion and reciprocity 'move in the same direction'. In part (a) we see that an increase in the links between pro-social players leads them to contribute more because of both guilt aversion and reciprocity. Similarly, in part (b), an increase in the links between pro-self players leads them to contribute less because of both guilt aversion and reciprocity. This allows us to make some predictions on the optimal network of information.

Corollary 2.1. *(a) Contributions are, ceteris paribus, weakly increasing in the number of links between pro-social players. (b) Contributions are, ceteris paribus, weakly decreasing in the number of links between pro-self players.*

We cannot theoretically tie down what happens if we increase the links between pro-social and pro-self players because this depends on the relative weight of guilt aversion and reciprocity. For instance, suppose we add a link between a pro-social player i and a pro-self player j. Player i will revise upwards her second-order beliefs and so be inclined to contribute more because of guilt aversion. But she will revise down her first-order beliefs and so be inclined to contribute less because of reciprocity. Similarly, player j will revise down her second-order beliefs and so be inclined to contribute less but revise up her first-order beliefs and so be inclined to contribute more.

Picking apart what happens in the case of pro-social to pro-self links is, therefore, an empirical matter. We will explore this with our experiment. We finish, however, by highlighting that Corollary 2.1 is sufficient to make interesting predictions. It shows that the optimal network of information will be heavily dependent on the SVO distribution in the group. If every player in the group is pro-social then the Complete network is predicted to be optimal. If every player in the group is pro-self then the Empty network is optimal. We can also say that it is optimal in the Star network to have a manager who is pro-social. All of these predictions can be tested in the laboratory.

2.4 Experiment design

In our experiment we set $m = 0.4$, $n = 4$ and $E = 20$. Thus, subjects were split into groups of four, given an endowment of 20 and faced a marginal per capita return of 0.4. The parameters selected follows the study by Fehr and Gächter [2000b]. As their study were replicated by many researches, we believe that if we are consistent with them, it will enable our results to be compared with those of other scholars. On the other hand, such parameter setting is relatively mature after years of experiments. Therefore, it can also be helpful for our research. We had four treatments corresponding to the four different networks of information in Figure 2.1: Empty (0 links), Pair (2 links), Star network (3 links) and Complete (6 links).

The timing of an experimental session was as follows: (a) Subjects were randomly assigned to terminals in the lab. (b) Subjects received instructions about the public good game and how they would be paid based on points earned. The instructions (available in the appendix) used neutral language around 'contributing points to a Group Account'. (c) Subjects were then told that 'Before we begin the group project and the main part of the experiment we will do a preliminary task'. They were given instructions for the SVO slider task [Murphy et al., 2011]. (d) Once they had done the slider task they were classified as pro-self if they had a slider angle less than 22.45° and classified as pro-social otherwise. (e) Subjects were told that:

> The task you have just performed is a standard psychological test used to classify people as either pro-self or pro-social. A person's behaviour is pro-social if it benefits others while it is pro-self if it benefits the person themselves. Based on the choices made, you and everyone else in your group has been classified as either pro-social or pro-self. Note: The classification of you and others in your group may be shared anonymously amongst the group members in the remainder of the experiment. You will be told if and when this happens.

(f) Subjects were asked to say whether they thought they would be classified as pro-social or pro-self. (g) Subjects were then told their SVO type and depending on the treatment told the SVO of others in the group. (h) Subjects played 10 rounds of the public good game in fixed groups. At the end of each round subjects were given feedback on the contribution of each group member. (i) After 10 rounds subjects were randomly re-matched into another group to play for a further 10 rounds in fixed groups. (j) Subjects completed a short questionnaire and were paid based on cumulative earnings over the 20 rounds.

Our SVO classification is standard in the literature and typically gives, as we wanted, around a 50-50 split between pro-self and pro-social individuals. We were careful to avoid any deception in the experiment, particularly in terms of sharing information about SVO. It is important that subjects completed the SVO before knowing it will be told to others - otherwise they may have behaved strategically when completing the SVO. But subsequently they were told that information about their SVO may be shared with others in the group and they had the option to leave the experiment at that point. Crucially we kept the experimental procedure as similar as possible across all treatments. For example, subjects in the Empty network were still exposed to the possibility that information about their SVO may be told to others.

We clarify that we did not elicit beliefs during the public good game. Our primary reason for not doing so was a concern about influencing behaviour (potentially though an experimenter demand effect). General issues around eliciting beliefs are discussed by Schotter and Trevino [2014] and Schlag et al. [2015]. The particular issue in our experiment is that if we give subjects information about SVO and then ask then to predict contributions it could prime or induce subjects to connect SVO with beliefs in a way they would not naturally have done. Hence, if we observe an effect of information, we would not have known whether that was induced by the elicitation of beliefs. We, thus, adopted a black-box approach in which we focus on the more fundamental question of whether the network of information influences contributions.

The experiment was run on computers using o-Tree [Chen et al., 2016]. It was conducted in the lab of the [double blind] with subjects recruited by the [double blind] system from the subject pool of the university. In total, 192 participants took part in the experiment (58.9% female).[2]

2.4.1 Experiment Hypotheses

In analysing the experimental results we will focus in particular on three hypotheses that directly follow from Corollary 2.1.

Hypothesis 2.1. *Average contributions are increasing in the number of pro-social to pro-social links.*

Hypothesis 2.2. *Average contributions are decreasing in the number of pro-self to pro-self links.*

Hypothesis 2.3. *Average contributions are higher in the Star network if there is a pro-social manager than a pro-self manager.*

Hypotheses 2.1 and 2.2 would imply that the effect of network structure on overall contributions is heavily dependent on the proportion of pro-social and pro-self players in the group. For instance, in the Complete network the number of pro-social to pro-social or pro-self to pro-self links can vary from 0 to 6 depending on the composition of the group. Applying our hypotheses we would expect contributions to be higher in a Complete network with 6 pro-social to pro-social links than the Empty network (with no links of any kind) to a Complete network with 6 pro-self to pro-self links.

Our model also allows us to draw testable predictions on the relationship between individual SVO and contribution to the group project. In specific, based

[2] 83.3% of them were undergraduate students and 16.7% masters students. All the participants had the experience of taking part in an economic experiment, where 37.5% of them had participated less than 5 times, 53.1% had participated 5 to 10 times and 9.4% more than 10 times.

on Proposition 2.2 in light of guilt aversion, we have the following hypothesis.

Hypothesis 2.4. *Individuals who are classified as pro-social (pro-self) contribute more (less) as the number of connected links increases.*

2.5 Results

In providing our experimental results we first examine the relationship between SVO and contributions to the public good. Next we test our hypotheses on the influence of network structure for individual contributions. Finally, we finish with an analysis of the Star network and the role of managers.

Table 2.1 provides an overview of key descriptive statistics across treatments and Figure 2.3 plots the distribution of SVO angle by treatments. Overall, there is no significant difference in the distribution of SVO types, own perception of SVO and other socio-economic variables across the four treatments. We do highlight, however, because it will be important for the interpretation of our results, that the absolute number of pro-socials is highest in the Complete network and lowest in the Empty network.

Table 2.1: Descriptive statistics

		Treatment				
	Overall	Complete	Star	Pair	Empty	*p*
Slider angle	17.0	20.0	17.0	15.2	15.9	*.355*
Pro-social (%)	38.5	50.0	39.6	33.3	31.2	*.385*
Own belief pro-social (%)	46.4	52.1	41.7	39.6	52.1	*.591*
Male (%)	41.4	43.8	39.6	37.5	43.8	*.936*
Age	20.7	20.3	20.9	21.0	20.5	*.262*
Undergraduate (%)	83.3	89.6	83.3	70.8	89.6	*.340*
Has lab experience (%)	62.5	54.2	58.3	70.8	66.7	*.478*
Single child (%)	48.4	43.8	50.0	47.9	52.1	*.909*
Observations	192	48	48	48	48	

Notes: The table reports average descriptive statistics of participants in each treatment. $p - value$ at the end of each row is derived from Kruskal–Wallis (KW) test for differences across treatments.

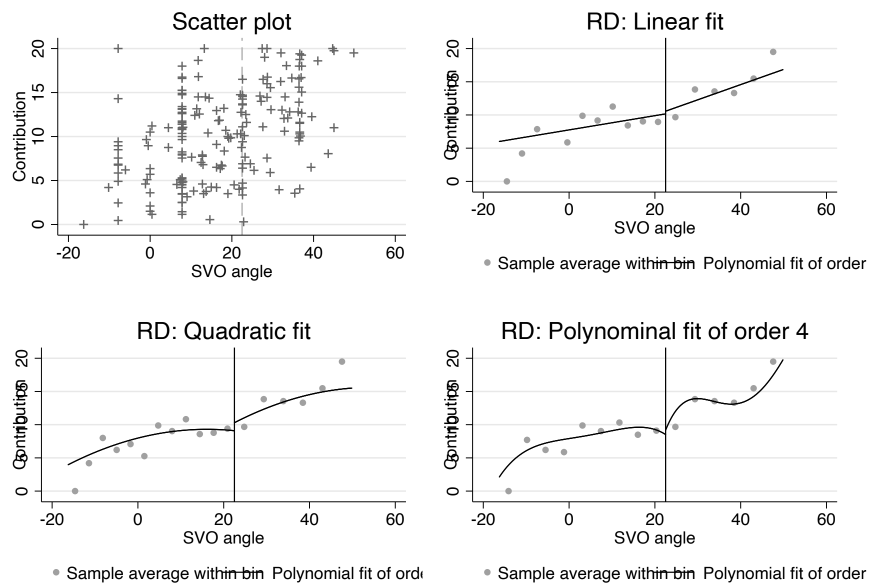


Figure 2.2: Contribution by SVO slider angle

We first look at individual contribution by SVO angle. Figure 2.2 reports scatter-plots of individual contribution against her SVO angle. It can be seen from the top left panel of there is an increasing trend between individual contribution and pro-sociality, as represented by higher SVO angle. Participants with SVO angle higher than 22.45° are classified as pro-social. To examine the pure effect of the classification, we used regression discontinuity (RD) analysis with a threshold of 22.45°. It can be seen from the RD plots (top right panel and bottom two panels, vary in Polynominal orders of fitness), there is a slightly jump at the threshold. Individuals who are of similar SVO angle close to 22.45° are slightly more likely to contribute when they are classified as pro-social. However this effect is statistically insignificant at any conventional level and bin width, so we found little evidence in the pure effect of labeling.

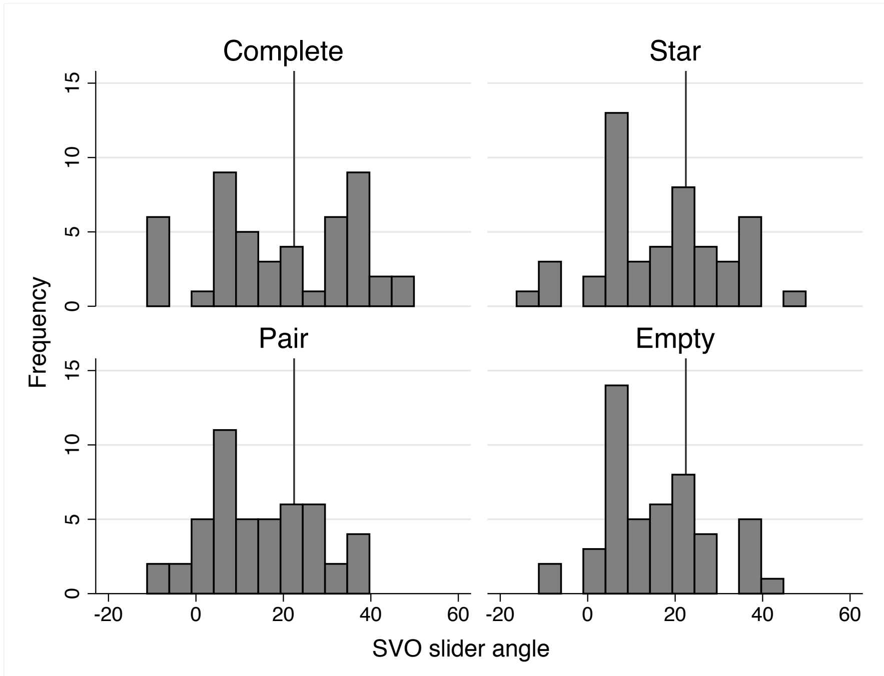


Figure 2.3: Distribution of SVO slider angle by treatments

2.5.1 SVO and individual contribution

Table 2.2 details the average contribution of pro-social and pro-self players in each treatment. We also report Mann-Whitney tests comparing pro-socials to pro-selfs using the group as the unit of observation. The effect size measures the probability a pro-social player contributes more than a pro-self player. One crucial thing to observe from Table 2.2 is that in the Empty network we find no significant difference in contribution between pro-socials and pro-selfs. Indeed, the median contribution of pro-socials is below that of pro-selfs. In principle, therefore, the assignment of pro-self and pro-social conveys no useful information.

Table 2.2: Contribution by treatment and SVO type

		SVO types			
Treatment		Pro-social		Pro-self	*Effect size*
Complete	*Mean*	14.3	$>^{***}$	8.8	*.82*
	Median	*13.8*		*8.7*	
Star	*Mean*	12.0	$>^{***}$	7.0	*.76*
	Median	*12.5*		*6.8*	
Pair	*Mean*	13.1	$>^{***}$	8.7	*.73*
	Median	*15.2*		*7.7*	
Empty	*Mean*	10.3	$\sim$	9.8	*.54*
	Median	*9.4*		*9.9*	

This table reports the mean and median contribution by SVO types and treatment. Unit observation is the individual average contribution across a session, each treatment has 48 observations. Effect-size and p-values are derived from Mann-whitney tests. In this and in the following tables, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Whether it is 'noise' or not, in the other three treatments you can see that there is a highly significant (in both statistical and economic terms) difference in contributions between pro-socials and pro-selfs. Recall that in all treatments subjects are informed of their own SVO. The difference between the Empty and other treatments provides, therefore, strong evidence that *sharing information* on SVO makes a critical difference. In particular you can see that pro-socials contribute more in the other three treatments than they do the Empty network while pro-selfs contribute less. This evidence, which we investigate more in the following, is consistent with Hypothesis 4 and points towards a significant role of guilt aversion.

Result 2.1. *Pro-social subjects contribute more than pro-self subjects. However the effects are mainly driven by treatments where SVO information is observable (i.e. in Complete, Star and Pair network). In the Empty network there is no significant differences between the contributions of pro-social and pro-self subjects.*

To reinforce result 2.1 and the apparent role of guilt aversion and second order beliefs, in Figure 2.4 we plot the average contribution of pro-social and pro-self subjects across treatments and the 20 rounds (with a restart at round 11). Here we can see that the stark difference between the Empty network and other three networks is apparent from round 1 on-wards. While pro-socials contribute marginally more than pro-selfs in round 1 and 11 of the Empty network the difference is not statistically significant ($p > 0.1$, Mann Whitney test) and considerably less than in all of the other three networks. Moreover, the gap between pro-social and pro-self subjects appears relatively stable across time in all networks (with the exception of an end-game effect in the Pair network for pro-socials).

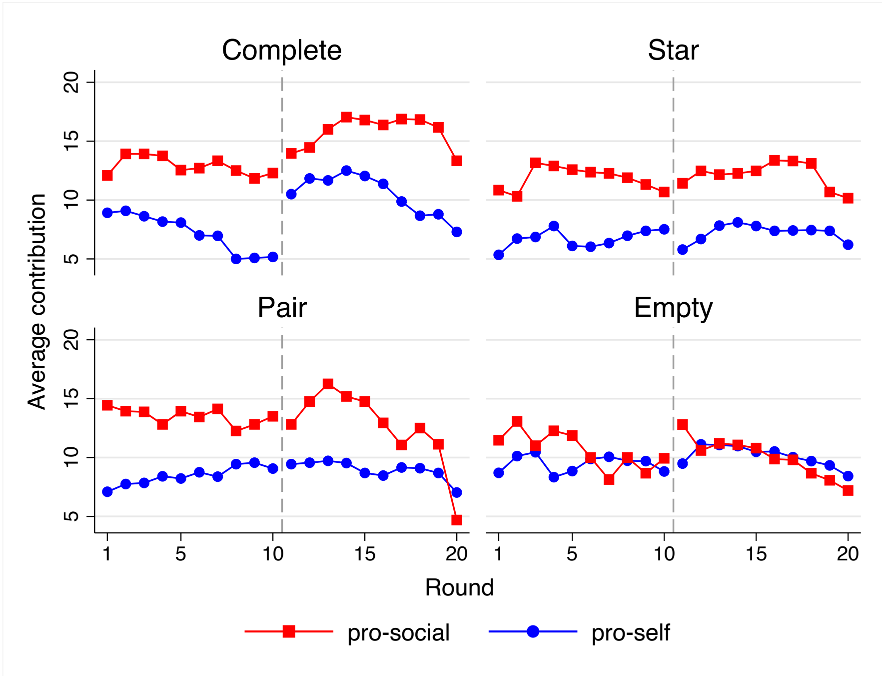


Figure 2.4: Dynamic individual contributions by SVO types and treatments

In Table 2.3 we report the results of a linear regression with individual contribution (averaged over 10 rounds) as the dependent variable. Independent variables include dummy variables for treatment, a dummy variable if the subject is pro-social, and dummy variables for the product of treatment and pro-social. In specification (2) we control for the rematch at round 11 and individual demographics including gender, age, university degree and prior lab experience. The baseline for comparison is a pro-self subject in the Empty network. To control for multiple testing across treatment and SVO type, we also report the Sharpened False Discovery Rate (FDR) q-values [Anderson, 2008] in the brackets in specification 1.

The statistically insignificant coefficient on 'pro-social' in Table 2.3 is consistent with pro-selfs and pro-socials contributing the same in the Empty network (Result 2.1). The statistically negative coefficient on 'Star' suggests that pro-selfs contribute

Table 2.3: Linear regression of individual contribution on treatment

Dependent variable: Average contribution across 10 rounds	(1)	(2)
Complete	-0.957	-0.717
	(1.134)	(1.134)
	[0.250]	
Star	-2.833***	-2.540**
	(1.040)	(1.100)
	[0.014]	
Pair	-1.094	-0.952
	(0.940)	(0.928)
	[0.173]	
Pro-social	0.535	0.0469
	(1.148)	(1.154)
	[0.379]	
Complete × Pro-social	4.969***	4.770***
	(1.673)	(1.622)
	[0.014]	
Star × Pro-social	4.496***	4.350***
	(1.554)	(1.573)
	[0.014]	
Pair × Pro-social	3.830**	4.356**
	(1.763)	(1.738)
	[0.031]	
Rematch		0.941
		(0.589)
Constant	9.788***	-0.003
	(0.705)	(4.102)
Control for individual characteristics	No	Yes
Observations	384	384
R-squared	0.144	0.169

Linear regression with bootstrapped standard errors. Unit observation is average individual contribution across 10 rounds, before and after restart. In specification (2), we control for restart and individual demographic characteristics including: gender, age, number of siblings, university degree and prior laboratory experiences. Sharpened False Discovery Rate (FDR) q-values are reported in the bracket to correct for multiple testing.[3]

less in the Star network than they do in Empty network. While the coefficient is negative for 'Complete' and 'Pair' it is not significantly so. There is, therefore, no compelling evidence that pro-selfs contribute less in these two networks than they do in the Empty network. By contrast, the coefficients on 'treatment × pro-social' are highly statistically significant for all three treatments. This is evidence that pro-socials contribute more when their SVO is observed by others.

Result 2.2. *Pro-social subjects contribute significantly more when their SVO is observed (Complete, Star and Pair networks) than when it is not (Empty network). Pro-self subjects contribute significantly less in the Star network than Empty network but there is no evidence they contribute significantly less in the Complete or Pair networks.*

2.5.2 Contributions and network structure

Result 2.2 points to an interesting asymmetry in which pro-social subjects increase their contribution relatively a lot when SVO is observed (by around 4.5) while pro-selfs do not decrease their contribution by as much (around 0.7 to 2.5 depending on the treatments). This effect has to be counter-weighed by the fact that more subjects were classified as pro-self than pro-social. Recall from Table 2.1 that around 39% of subjects were pro-social with the number highest in the Complete network (50%).

Figure 2.5 details average group contributions by treatments across the 20 rounds. Overall, you can see that contributions are relatively high in the Complete network (average 46.33 tokens) and relatively low in the Star network (35.79 tokens) compared to the Empty and Pair networks (39.82 and 40.60 tokens respectively). The difference between the Star network and Empty network is statistically significant in both rounds 1-10 and rounds 11-20 ($p < 0.01$, Wilcoxon test with groups as the unit of observation). The difference between the Complete and Empty network is statistically significant in rounds 11-20 ($p < 0.01$) but not rounds 1-10 ($p > 0.1$).

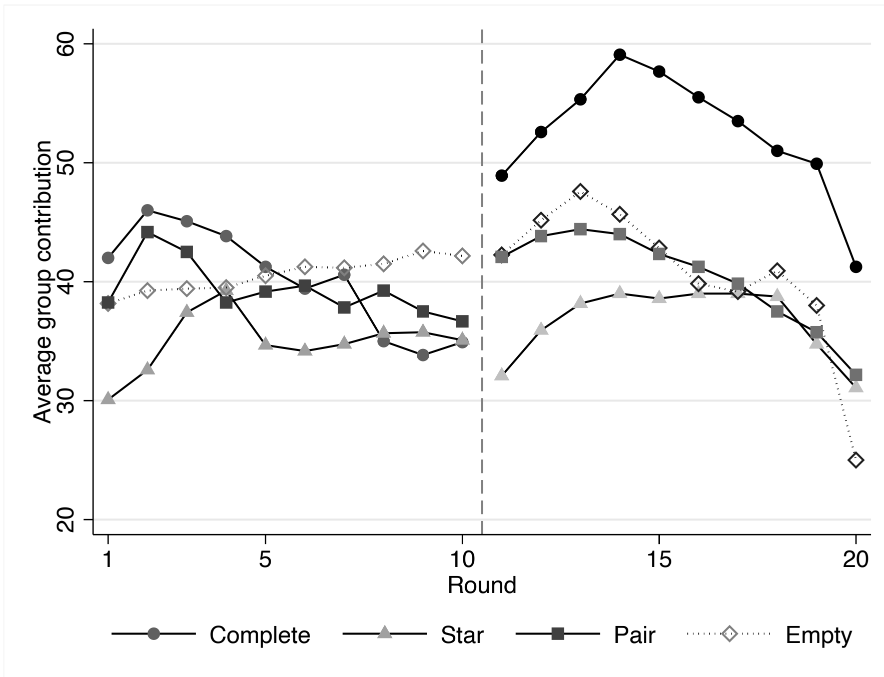


Figure 2.5: Dynamic group contributions by treatments

These treatment differences are apparent in round 1 and round 11. Specifically, in rounds 1 and 11 contributions are highest in the Complete network (round 1: mean = 42.0 tokens, SD = 12.6; round 11: mean = 48.9, SD = 11.8). They are lowest in the Star network (round 1: mean = 30.1, SD = 13.4; round 11: mean = 32.1, SD = 13.7). The Empty and Pair networks are intermediate (round 1: mean = 38.2 & 38.1, SD = 8.0 & 17.7; round 11: mean = 2 & 42, SD = 10.7 & 13.1). Mann-Whitney test results suggest statistically significant differences in group contributions between Star and the other three networks in rounds 1 and 11, and those between Complete and Empty/Pair in round 11.[4]

To help interpret these treatment differences we now turn our attention to Hypotheses 2.1 and 2.2. In Figure 2.6 we plot average contributions in the group as a function of the number of pro-social links in the group. In the Star network we see that average contributions are unambiguously higher the more pro-social links in the network. This is consistent with Hypothesis 2.1. The picture is similar in the Complete and Pair network although slightly less clear-cut (when looking at 6 in the Complete network and 1 versus 2 in the Pair network).

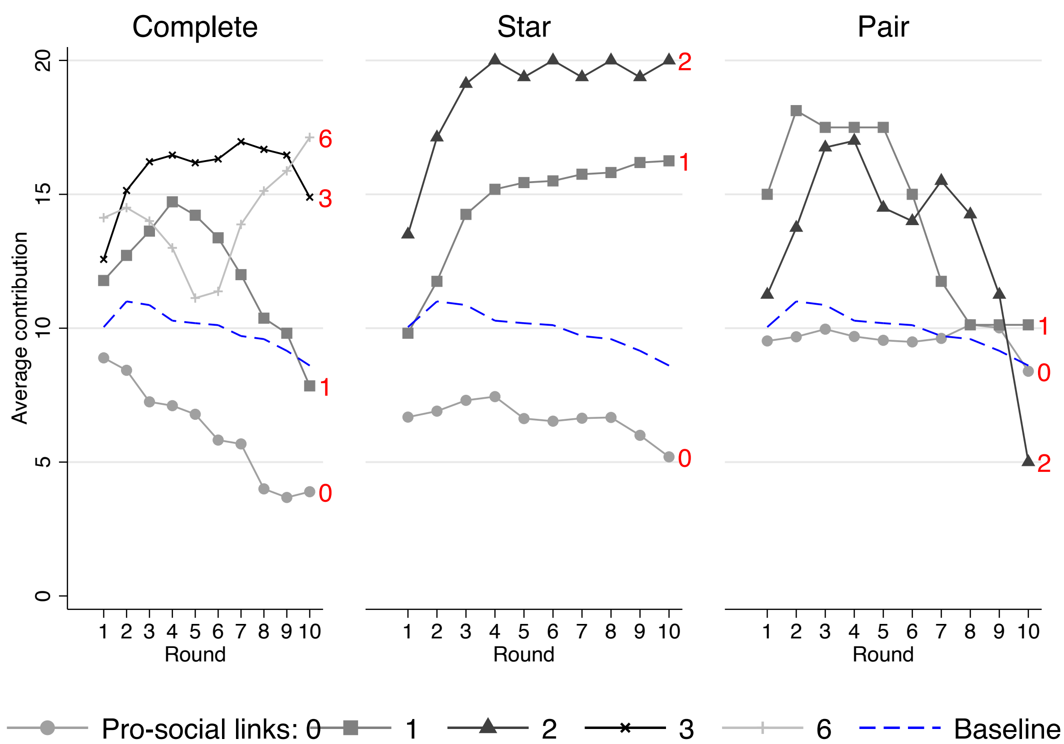


Figure 2.6: Contribution by numbers of pro-social links and treatments

[4]Mann-Whitney test results: Star and Pair in round 1 & 11: $p = 0.0399$ and $p = 0.0008$; Star and Empty in round 1 & 11: $p = 0.0015$ and $p = 0.0027$; Star and Complete in round 1 & 11: $p = 0.0001$ and $p = 0.000$; Complete and Empty in round 11: $p = 0.0039$; Complete and Pair in round 11: $p = 0.0160$.

In Table 2.4 we report the results of a regression analysis with individual contribution as the dependent variable. As independent variables we include whether or not the subject was classified as pro-social, the number of pro-socials the subject was linked to, whether the treatment was non-empty and the interaction term between pro-social and whether the treatment was non-empty. In interpretation we can think of the Empty network as the baseline.

The positive but insignificant coefficient on Pro-social reinforces Result 2.1. The negative coefficient on 'Non-Empty Treatment' is consistent with pro-selfs contributing less when their SVO is observed. The positive coefficient on 'Pro-social × Non-Empty' suggests that pro-socials contribute significantly more when their SVO is observed. Crucial for our purposes are the effect of pro-social links. Consistent with Hypothesis 2.1 you can see that an individual's contribution is strongly increasing in the number of pro-social links. Which implies that, due to linear combination of pro-social and pro-self links within each treatment, contributions are strongly decreasing in the number of pro-self links (consistent with Hypothesis 2.2). Across the three non-Empty treatments, the expected and actual proportion of pro-social to pro-social links is highest in Complete network (expected: 25%; real: 28.5%), followed by Star (12.5%; 11.1%) and Pair (9.375%; 8.3%), which helps to explain the differences in group contributions that we observed in Figure 2.5.

Result 2.3. *Contributions are significantly increasing in the number of pro-social links in the network. They are significantly decreasing in the number of pro-self links.*

Table 2.4: Linear regression of individual contribution on SVO information

Dependent variable: Individual contribution				
	Round 1	Round 1-10	Round 11	Round 11-20
Pro-social	2.770	1.176	3.315*	-0.105
	(2.076)	(1.123)	(1.698)	(2.013)
Non-Empty treatment	-2.515*	-3.213***	-2.281*	-3.720***
	(1.341)	(1.134)	(1.349)	(1.249)
Pro-social×Non-Empty	2.236	3.575**	0.882	4.823**
	(2.368)	(1.436)	(2.054)	(2.063)
Number of links with pro-social	1.195***	1.806***	1.687***	3.115***
	(0.432)	(0.502)	(0.538)	(0.441)
Constant	8.697***	9.464***	9.485***	10.11***
	(1.204)	(0.769)	(1.034)	(1.083)
Control for individual characteristics	Yes	Yes	Yes	Yes
Observations	192	192	192	192
R-squared	0.165	0.197	0.138	0.265

Linear regression with bootstrapped standard errors. Unit of observation for Round 1 and Round 11 is individual contribution in that round; for Round 1-10 and Round 11-20 is average individual contribution for the ten rounds. In all specifications, individual characteristics including gender, age, number of siblings, university degree and prior laboratory experiences are controlled.

2.5.3 Star Network

We conclude the analysis by looking at the Star network. Recall that in this network one subject, who we shall call the 'manager', is informed of the SVO of all group members. In Figure 2.7 we plot the average contribution of the manager and the average contributions of others (Non-manager) in the group split by whether the manager is pro-social or pro-self. You can see that contributions of both the manager and non-managers more than double if the manager is pro-social compared to pro-self ($p < 0.05$ Mann Whitney test). This is strong evidence in support of Hypothesis 2.3.

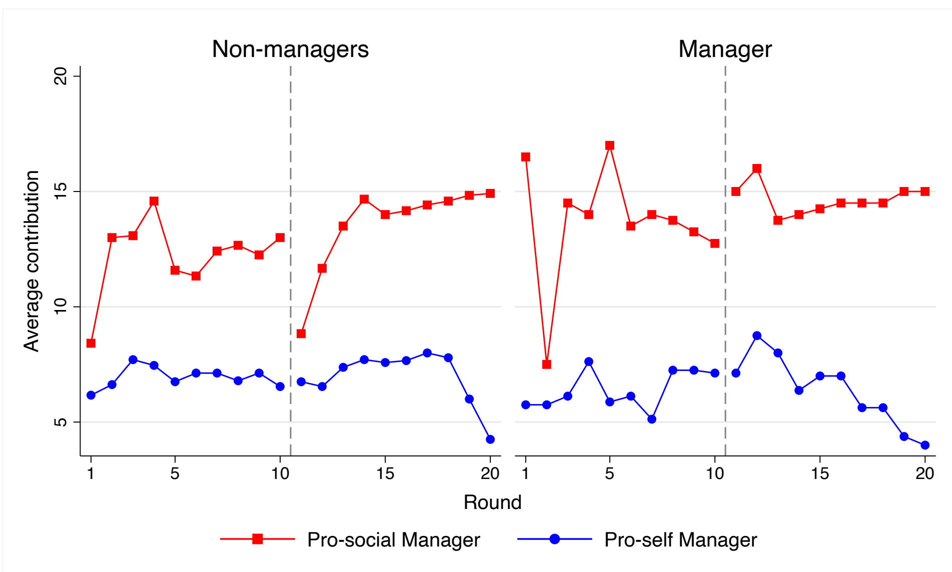


Figure 2.7: Contribution by manager's SVO type

Result 2.4. *Managers and non-managers contribute significantly more in the Star network if the manager is pro-social compared to if the manager is pro-self.*

It is particularly interesting to note the dynamic effect in Figure 2.7 whereby the contributions of non-managers increase over the first few periods if the manager is pro-social. This suggests that a pro-social manager may serve as a leader in the group in the sense of Andreoni and Petrie [2004]. In other words, they keep contributions 'high' in order to pull up others. If so, then pro-social managers may not only lead to higher contributions but might also be able to reverse previous coordination failure [Brandts et al., 2015]. This is an issue that could be analysed in future work.

2.6 Conclusion

In this experiment we have explored whether information about social value orientation (SVO) influences contributions to a public good. We did so by comparing an Empty network, in which individuals only know own SVO, with a Pair, Star and Complete network, in which individuals get to know each others SVO. We hypothesized that information about SVO would influence contributions through the effect it has on first and second order beliefs. We found a strong effect of information, consistent with both reciprocity and guilt aversion [Dufwenberg et al., 2011]. In particular, consistent with guilt aversion, pro-socials contributed more and pro-selfs contributed less when their SVO was observed by others. And, consistent with reciprocity, contributions were increasing in the number of pro-social links and decreasing in the number of pro-self links.

Our results provide further evidence that making information on psychological traits known within the group can influence group behavior [Drouvelis and Georgantzis, 2019]. Crucially, this happens despite SVO being a weak predictor of contributions in a setting where information is private. In other words, group members are strongly influenced by what is essentially noisy information. In our experimental analysis we observed a slight asymmetry towards information having a positive effect. In particular, pro-social individuals in the group tended to have a bigger positive effect than pro-self individuals had a negative effect. Hence, information was a net positive. We note, however, that Drouvelis and Georgantzis [2019] obtain a contrasting result in a two player public good game where information was a net negative. The primary reason we observe a net positive effect is that those subjects who are pro-social significantly increase their contribution.

It is interesting to question what type of network structure optimizes contributions. This question is posed in a setting where we cannot control the number of pro-socials in the group merely the network of information. As one might expect, if all group members are pro-social then our results suggest the optimal network would be the Complete network. Similarly, if all group members are pro-self the optimal network would be the Empty network. But what about the more relevant middle ground? Our results suggest that the optimal network is the Star network *provided* the manager is pro-social. Specifically, with the Empty network average contributions are around 50% of endowment and in the Complete network around 60% of endowment, while in the Star network with a pro-social manager they rise to around 70% of the endowment. Moreover, in the Star network with a pro-social manager contributions remain stable over time.

In our setting the 'manager' is merely the member of the group who is most informed. It is unclear, therefore, to what extent the positive effect of a pro-social manager is driven by the effect it has on the manager, or the indirect effect on others of having a pro-social 'focal point'. We know that contributions to a public good are highly dependent on the dynamics of reciprocity [Croson, 2007, Gächter et al., 2017] and so small differences can amplify into bigger effects. Eckel et al.

[2010] consider a public good game with a Star network in terms of observing each others contributions. Interestingly, they find that a high status ‘manager’ is more likely to be ‘followed’. Potentially, therefore, a pro-social manager is more likely to be followed, creating a significant indirect effect of having such a manager. These are issues that could be explored in future experiments. We finish by noting that our findings reinforce the evidence that a pro-social or servant leader is good for cooperation within groups [Gillet et al., 2011, Chiaburu et al., 2014].

Chapter 3

Pro-social norms and feedback on own social value orientation

[0]Work joint with Edward Cartwright (De Montfort University, UK), Lian Xue (Wuhan University, China)

[0]Each researchers' contribution: the idea, original experimental design and coding, experiment conducting, major parts of analysis by Yidan; theory, model suggested and help by Prof. Cartwright; Dr Xue helped data analysis as well.

[0]Funding for this project was provided by Department of Economics and Marketing, De Montfort University, Leicester, UK.

3.1 Introduction

There is extensive evidence that a large proportion of individuals are willing to contribute to the provision of public goods. Participants on average contribute 40% to 60% of endowments games, this indicates that many people will contribute in public goods games voluntarily, though there is no apparent and direct interests given to them [Chaudhuri, 2011]. Various explanations exist in the literature for such pro-social behaviour, ranging from confusion to social preferences to social norms [e.g. Burton-Chellew et al., 2016, Fehr and Schurtenberger, 2018, Goeschl and Lohse, 2018, Henrich and Muthukrishna, 2021]. Carefully distinguishing between these different explanations is vital to improve our understanding of pro-social behaviour and thereby design appropriate behaviour change interventions. In this experiment we propose and analyse one potential way to distinguish between two broad categories of explanation for pro-social behaviour - belief based preferences versus internalized norms based on empirical expectations. The approach we propose is based on giving individuals private information on their own social value orientation. We argue that this private information should not influence contributions if contributions are driven by belief based preferences but should if they are driven by internalized norms. We report the results of an experiment where private information about social value orientation was given to subjects. This private information is shown to have no impact on pro-social behaviour.

In motivating our approach let us first provide a brief overview of social value orientation (SVO). SVO is a measure of an individual's willingness to share money with an anonymous other person [Van Lange, 1999]. It can be used to make a distinction between pro-social individuals who show a willingness to share money and pro-selfs (or individualists) who do not [Murphy et al., 2011].[1] A large number of studies have shown a positive relationship between an individuals SVO and willingness to contribute to a public good [e.g. Balliet et al., 2009, Pletzer et al., 2018] (exception see [Renkewitz et al., 2011]). In a companion paper we demonstrate that public information on SVO has a predictable and significant effect on contributions [Cartwright, 2019]. Particularly relevant for our purposes here, we find that pro-social individuals contribute more and pro-self individuals contribute less if there SVO is known by another member of the group. Manipulating public information on SVO can, therefore, promote pro-social behaviour. In this experiment we turn our attention to private information on SVO.

In exploring the potential consequences of private information we compare two broad categories of behavioural model. Various models in psychological game theory posit that preferences and therefore behaviour are influenced by first, second and higher order beliefs on the contributions of others. For instance, contributions may be increasing in first order beliefs on the contributions of others because of reciprocity [Rabin, 1993, Dufwenberg and Kirchsteiger, 2004, Falk and Fischbacher, 2006, Smith, 2013, Dufwenberg and Patel, 2017]. Similarly, contributions may be increasing in second order beliefs because of guilt aversion [Battigalli and Dufwenberg, 2007, Dufwenberg et al., 2011, Caria and Fafchamps, 2017]. Crucially, if behaviour is driven by beliefs about others then private information should not influence behaviour because 'others' are not exposed to that information. We argue, therefore, that giving individuals information about their own SVO should not influence contributions to a public good if those contributions are driven by belief based preferences.

[1] More types are possible to distinguish but most individuals fit in the pro-social or individualistic category.

An alternative approach is to view contributions as influenced by social norms [Villeval, 2020]. Broadly speaking we can define a social norm as a behaviour that someone is expected to follow in a particular situation [Lapinski and Rimal, 2005] or ought to follow [Fehr and Schurtenberger, 2018]. If a norm is internalized then an individual may feel guilt or shame in deviating from that norm creating an internalized incentive to behave in a particular way [Bicchieri et al., 2018, Kelly, 2020]. Cooperation is one norm that may be internalized [Gintis et al., 2004, Burger et al., 2009, Villatoro et al., 2015, Gavrilets and Richerson, 2017, House, 2018]. Crucially, we would argue that someone labelled pro-social would be expected to contribute more than someone labelled pro-self. This may follow from either of two related possibilities. First, that individuals labelled pro-self feel less internalized incentive to contribute. Second, that individuals labelled pro-social feel a heightened incentive to contribute. Either would imply that private information about SVO can be expected to have an impact on contributions.

Exposure to private information on SVO provides, therefore, an interesting way of differentiating belief based and internalized norm models of cooperation [Cartwright, 2019]. We report the results of an experiment designed to test if private information on SVO has an impact on behaviour in a one-shot public goods game. In the experiment subjects undertake the slider measure test of SVO [Murphy et al., 2011]. We then use a variant of the Fischbacher et al. [2001] protocol to elicit unconditional and conditional contributions to the public good. We compare three treatments: (i) no information is given on SVO, (ii) subjects are told their binary classification into pro-self or pro-social, (iii) subjects are given a sliding measure of SVO ranging from very pro-social to very pro-self. Our main result is that we find no difference, either economically or statistically, in contributions across the three treatments. This, we argue, provides support in favor of belief based models of behaviour as compared to models of internalized norms. In forwarding this argument we recognise that there are many alternative interpretations of our findings, such as subjects simply ignoring SVO or norms not being influenced by SVO. In the analysis and concluding discussion we explore and evaluate these alternative interpretations.

We proceed as follows: In Section 3.2 we introduce a public good game with leadership. In Section 3.3 we provide theoretical predictions. In Section 3.4 we discuss our experiment design and in Section 3.5 provide the results. In Section 3.6 we conclude. Additional material is provided in the appendix.

3.2 Public Goods Game with Leadership

There is a set $N = \{1, ..., n\}$ of players. Each player is endowed with E units of a private good and has the opportunity to contribute to a public good. Let $c_i \in [0, E]$ denote the contribution of player $i \in N$. The material payoff of player $i \in N$ is given by

$$u_i = E - c_i + m \sum_{i=1}^{n} c_i$$

where $1/n < m < 1$ is the marginal per-capita return from the public good. Given that $m > 1/n$, total payoff in the group is maximised if every player in the group contributes their full endowment E.

We consider a game with leadership by example in which one player in the group, the leader, makes a contribution before the other $n-1$ players, the followers [Sahin et al., 2015, Gächter and Renner, 2018]. Without loss of generality we assume that player 1 is the leader. Followers can make their contribution conditional on that of the leader. A strategy for a follower consists, therefore, of a function mapping from the leader contribution to own contribution.

The novelty in our work is to allow that, before the public good game, each player $i \in N$ receives feedback on a test to measure SVO. SVO is typically measured by the angle of cooperation [Murphy et al., 2011]. While this can take any value it almost always lies in the range of $-16.26°$, indicating the individual only cares about own payoff, to $61.39°$, indicating a willingness to share payoffs equally. Let v_i denote the SVO angle for player i.

We consider the possibility that SVO is revealed on a scale from very pro-self, $-16.26°$, to very pro-social, $61.39°$. Thus, let $v_i^c = v_i$ if $16.26 \leq v_i \leq 61.39$, $v_i^c = -16.26$ if $v_i < 16.26$ and $v_i^c = 61.39$ otherwise. We also consider the possibility that SVO is revealed as a binary pro-self or pro-social indicator. Let $v_i^b = 0$ if $v_i \leq 22.45$, and $v_i^b = 45$ otherwise. This gives three scenarios to compare: (i) *Scale measure*, each player $i \in N$ is privately told v_i^c, (ii) *Binary measure*, each player $i \in N$ is privately told v_i^b, and (iii) *No information*, no player $i \in N$ is given any information about v_i.

3.3 Theoretical Predictions

In this section we will draw a comparison between two broad ways of modelling behaviour in a public goods game - belief based versus internalised norm based. To illustrate the argument we will focus on guilt aversion because that provides a simple and transparent framework to consider. The argument, however, naturally extends, as we discuss below, to other frameworks such as reciprocity.

Consider the simple model of guilt aversion proposed by Battigalli and Dufwenberg [2007, 2009] and applied to public goods game by Dufwenberg et al. [2011]. Let b_{ij} denote player i's first order belief on the contribution of player j. In our setting, b_{i1} is player i's belief about the contribution of the leader and b_{ij} for $j \neq 1$ is player i's belief about the contribution of follower j. Let a_{iji} denote player i's second order beliefs. Specifically, a_{iji} denote player i's beliefs about b_{ji}. With a slight abuse of notation, let

$$a_i = \frac{1}{n-1} \sum_{i=1, i \neq j}^{n} a_{iji} \tag{3.1}$$

denote the average second order belief of player i. In our setting, a_1 can be interpreted as the amount the leader expects others believe she will contribute. Similarly, a_i for $i > 1$ can be interpreted as the amount a follower expects others believe she will contribute, conditional on the leader contributing c_1.

Following Dufwenberg et al. (2011) we can write the utility of player i as

$$u_i^b = E - c_i + m \sum_{i=1}^{n} c_i - \gamma_i max\{0, a_i - c_i\}$$

where γ_i measures the strength of guilt aversion. In interpretation a player feels guilt if she contributes less than she believes others expected her to contribute. The interpretation is that by contributing less than expected she lowers the payoffs others expected and, hence, feels guilt from 'disappointing' others. Her utility, thus, depends on her second order belief. Moreover, if γ_i is sufficiently high it is optimal for her to contribute a_i and so contributions are increasing in second order belief.

Now suppose that player i is informed of own SVO either in the form v_i^c or v_i^b. Can we expect that to have a discernible impact on behaviour? Given that the information is private we argue that it cannot be expected to influence first or second order beliefs in any systematic way. It is possible that common knowledge everyone knows own SVO may alter framing which can influence beliefs [e.g. Ellingsen et al., 2012]. There is, however, no clear rationale for why information on SVO would alter framing or influence contributions in a particular direction. The only plausible mechanism we can put forward is that a player informed of own SVO extrapolates this information onto others. For instance, someone informed they are pro-self increases the probability they put on others being pro-self, which then influences beliefs. This belief, though, is not internally consistent given that some players will clearly be learning they are pro-social.

On the basis that private information should not influence beliefs, or strength of guilt aversion, we put forward our first hypothesis.

Hypothesis 3.1. *Belief based: Private information on SVO has no influence on contributions to the public good.*

Hypothesis 3.1 provides a clear prediction that is readily testable in the lab by comparing treatments with and without information on SVO.

We turn now to an internalised norm based approach [López-Pérez, 2010]. Let e_i denote player i's norm of the contribution 'somebody like her' should make. This norm could be empirical - what she expects others would do in this situation - or normative - what she thinks is the right thing to do in this situation. For player i the contribution e_1 captures her norm for leaders. For players $i > 1$ it captures her norm for followers conditional on the leader contributing c_1. The utility function of player i can be written

$$u_i^b = E - c_i + m \sum_{i=1}^{n} c_i - \delta_i max\{0, e_i - c_i\}$$

where δ_i measures the strength of guilt aversion. A player, thus, feels guilt if they contribute less than they believe they should have done.

In the internalized norm based approach there is straightforward mechanism whereby private information on SVO may influence contributions. For instance, suppose player i's norm is based on an empirical expectation of what 'most people do in this situation'. Let e_i^0 denote the mean contribution for someone with no information on SVO, and e_i^s and e_i^f the mean contribution for someone who is classified pro-social and pro-self, respectively. It would be consistent with the evidence, both from the previous literature and the results we will present here, to set $e_i^f < e_i^0 < e_i^s$. If, therefore, player i's norm is based on what others do then, for sufficiently large δ_i, her optimal contribution is higher if she is informed she is pro-social and lower if she is informed she is pro-self, compared to the baseline of no information. We obtain, therefore, a second hypothesis.

Hypothesis 3.2. *Internalized norm based: Private information on SVO influences contributions to the public good. Players labelled as pro-self contribute less and players labelled as pro-social contribute more than in a no information setting.*

Hypothesis 3.2 provides, again, a clear prediction that is readily testable in the lab by comparing treatments with and without information on SVO. It also provides a stark alternative to Hypothesis 3.1. Our experiment will test how well these two hypotheses perform. We briefly note that in our experiment we will control for what SVO the subject thinks they are. This is relevant for the internalized norm based approach because individuals tend to overestimate how similar they are to others [Fischbacher and Gächter, 2010]. Specifically, if i is pro-social we might expect that e_i^0 is close to e_i^s and if they are pro-self that e_i^0 is close to e_i^f. Private information may, therefore, have a bigger influence if an individual's SVO classification clashes with her prior belief. For instance, if individual i think she is pro-self but is told she is pro-social then we might expect a bigger effect on contributions than someone who expected to be classified pro-social.

Before we move onto the experiment, let us reiterate that the arguments presented above are not specific to the two formulations of guilt aversion proposed above. For instance, Hypothesis 3.1 readily extends to guilt from disapproval, as considered by Hauge [2016], because disapproval is based on what others believe. It also extends to psychological game theoretic models of reciprocity because those models assume payoffs and incentives depend on beliefs. Similarly, Hypothesis 3.2 extends to internalized norms of reciprocity. The distinction between Hypotheses 3.1 and 3.2 lies in whether private information can be expected to influence behaviour.

3.4 Experiment design

The experiment had three treatments, previewed in Section 3.2: no-information on SVO, scale measure of SVO and binary measure of SVO. Each subject was exposed to only one treatment. We, thus, obtain a between subject comparison on the influence of private information on SVO.

We highlight that all treatments were identical until the point where SVO was, or was not, revealed. Specifically, at the beginning of the experiment subjects were given the instructions to the public goods game. These instructions are available in the appendix. They were then told that before beginning the game they would do a preliminary task. At this point, SVO was measured using the slider test of Murphy et al. [2011]. In the test subjects are presented with six independent choice scenarios in which their choice determines the payoff for themself and an anonymous other person. Based on the choices made an SVO angle is calculated. Subjects were told:

> The task you have just performed is a standard psychological test used to classify people as either pro-self or pro-social. A person's behaviour is pro-social if it benefits others while it is pro-self if it benefits the person themselves. Based on the choices made, you and everyone else in your group has been classified as either pro-social or pro-self.

Subjects were then asked to indicate whether they thought they would be classified as pro-self or pro-social.

At this point in the experiment the three treatments diverge. In the No information treatment subjects proceeded to the public goods game with no feedback on their own SVO. In the Binary treatment every subject was told their classification as pro-social or pro-self before proceeding to the public goods game. An SVO angle of less than 22.45° was interpreted as pro-self and above 22.45° as pro-social. In the Scale treatment every subject was told their SVO classification on a scale rated from very pro-self (angle of −16.26°) to very pro-social (61.39°). They then proceeded to the public goods game. In the Binary and Scale treatments knowledge of SVO is private information, and commonly known to be private information.

We used a variant of the Fischbacher et al. [2001] protocol for measuring conditional cooperation proposed by Cartwright and Lovett [2014].[2] Subjects are split into groups of $n = 4$, endowed with $E = 20$ units of private money and have a marginal per-capita return from the public good of $m = 0.4$. Each subject is

[2]The original Fischbacher et al. [2001] protocol has 3 'first movers' and 1 follower in the group. The revised protocol has 1 leader and 3 followers. The leader follower protocol we use has shorter instructions and is arguably easier for subjects to understand. Cartwright and Lovett [2014] find no difference in behaviour across the two protocols.

first asked to make an unconditional contribution to the public good. This will be their contribution if they selected as the leader in the group. Each subject is then asked to fill in a contribution table that details the contribution they would make conditional on the contribution of the leader. Given that the contribution of the leader can be anything between 0 and 20 the contribution table has 21 entries. Once all four members of the group have made their choices, one member of the group is randomly selected to be the leader and outcomes are determined.

The experiment was conducted online using o-Tree [Chen et al., 2016] with subjects recruited on Prolific[3]. We recruited subjects who were resident in the UK and in employment.[4] Each subject was able to proceed through the tasks - SVO test, one-shot public goods game and questionnaire - at their own speed. The assignment of leader and calculation of payoffs was subsequently determined and subjects were paid. Subjects were paid a fixed fee for completing the experiment of £1.90 and a bonus payment, averaging £1.32, that was determined by points earned in the public goods game. Subjects were informed that the public goods game would be payoff incentivized. The experiment took 15-25 minutes for the vast majority of subjects.

In total there were 342 participants who completed the experiment across the three treatments. They were 68% female, 72% aged between 18 and 40, 59% in full-time employment, 58% have an undergraduate degree qualification and 61% have income above £30,000. We have 121 observations for the No info treatment, 120 for the Binary treatment and 101 for the Scale treatment. Individual characteristics including SVO are not statistically significantly differ across the three treatments

[3] https://www.prolific.co

[4] This restriction was motivated by a follow on questionnaire looking at cyber-security in the workplace.

(see Table A1 and Figure A1 in the Appendix for details). Note that because subjects complete the experiment independently of each other (and the group allocation is merely to determine a subject's payoff) we can treat each subject as an independent observation.

3.5 Results

In the analysis we will primarily focus on the comparison between pro-socials and pro-selfs (where the dividing line is a slider angle of 22.45°). This allows a straightforward direct comparison across the three treatments. In Table 3.1 we detail the number of subjects by type and treatment. The differences in proportion across treatments are not statistically significant ($p > 0.1$, proportions test).

Table 3.1: Number of subjects of each type and proportion who are pro-social by treatment.

Classification	Overall	No info	Binary	Scale
Pro-social	205	73	65	67
Pro-self	137	48	55	34
Pro-social (%)	60	60	54	66

We begin the analysis by looking at the unconditional contribution to the public good. In Figure 3.1 we plot the cumulative distribution of unconditional contribution by treatment and SVO type. Two key findings are readily apparent. First, subjects classified as pro-social contribute significantly more than those who are pro-self ($p < 0.01$ Mann-Whitney test for each of the three treatments). This is consistent with prior evidence [e.g. Balliet et al., 2009]. Second, and particularly important for our purposes, there is no evidence that the contribution of pro-socials or of pro-selfs differs across the three treatments ($p > 0.1$, Kruskal-Wallis test for both pro-socials and pro-selfs). The only sign of any difference comes in terms of pro-selfs where those in the No info treatment are essentially less likely to contribute 10 than those in the Binary and Scale treatment. This difference, however, is also not significant at conventional levels (0.25 versus 0.39, p = 0.094, proportions test). Overall, therefore, we see strong evidence in support of Hypothesis 3.1.

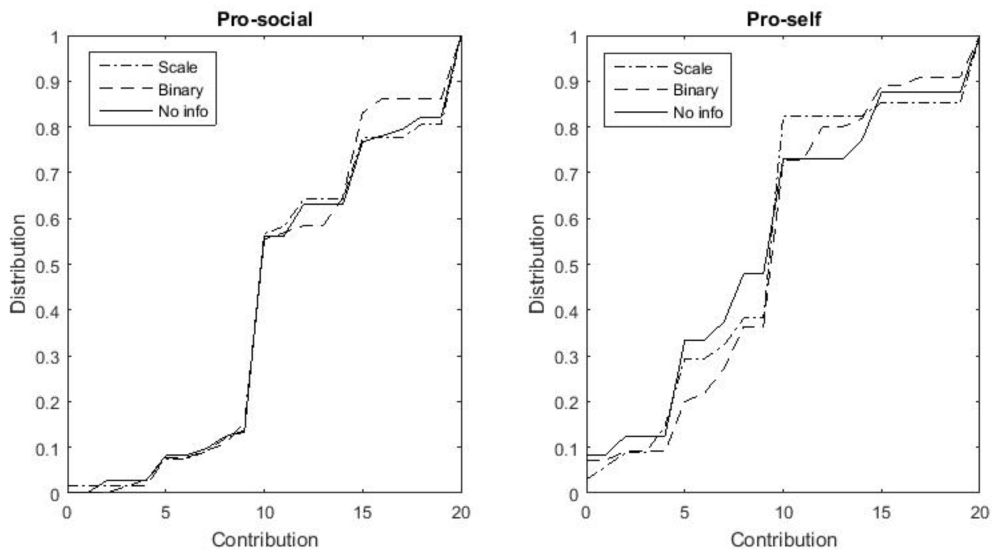


Figure 3.1: Cumulative distribution of unconditional contribution by SVO type and treatment

Table 3.2 reports the results of linear regressions with individual unconditional contribution as the dependent variable. Independent variables are treatment and treatment crossed with whether the subject is pro-social. We provide a regression with and without controls for demographics and the subjects belief of own SVO type. The baseline in both regressions is a pro-self subject in the No info treatment. The results reported in Table 3.2 reaffirm the two key findings identified above. There is evidence that pro-social subjects contribute more than pro-self subjects, although we note that the estimated effect reduces in magnitude once age and education are controlled for. There is no evidence of any effect of treatment, and information about SVO, on contributions.

Table 3.2: Linear regression on unconditional contribution

Dependent variable: Individual unconditional contribution		
	(1)	(2)
Binary	0.386	0.257
	(1.166)	(1.124)
Scale	0.134	0.253
	(1.461)	(1.466)
Pro-social	3.001***	1.985**
	(1.037)	(1.007)
Binary × Pro-social	-0.537	-0.161
	(1.458)	(1.397)
Scale× Pro-social	-0.128	-0.426
	(1.717)	(1.738)
Gender		0.440
		(0.645)
Age range		0.499**
		(0.196)
Employment status		0.144
		(0.204)
Income range		0.163
		(0.109)
Education level		0.325**
		(0.133)
SVO type belief		1.595***
		(0.418)
Constant	9.396***	2.934
	(0.885)	(1.857)
Observations	342	342
R-squared	0.072	0.133

Note: Linear regressions with bootstrapped standard errors. Dependent variables is individual unconditional contribution to the public goods game. Baseline treatment: no information. Bootstrapped standard errors in the parenthesis. *, **, and ***indicate significance at the 10%, 5%, and 1% levels, respectively.

Result 3.1. *There is no evidence that unconditional contributions to the public good are influenced by information on own SVO.*

Recall that subjects were asked whether they thought they were pro-social or pro-self before any information about SVO was revealed. Also recall, see Section 3.3, that it is particularly interesting to look at subjects whose beliefs are inconsistent with the SVO classification they received. In Table 3.3 we plot mean unconditional contribution by treatment and SVO classification distinguishing whether the subject's belief was consistent. In the case beliefs are consistent you can see that there is no evidence of any treatment effect. Interestingly, in the case of inconsistent beliefs there is some suggestion that contributions are higher with private information. These treatment differences are, however, not significant at conventional levels ($p = 0.051$ in Mann-Whitney test on all those classified as inconsistent and $p > 0.1$ for pairwise comparisons). Moreover, the notion that those classified as pro-self would increase their contribution is hard to reconcile with Hypothesis 3.2 and the internalized norm based approach. If anything, it would seem that 'inconsistency' is a trigger for subjects to contribute more irrespective of what the SVO information conveys.

Table 3.3: Number of subjects of each type and proportion who are pro-social by treatment.

Belief	Classification	No.	Mean contribution		
			No info	Scale	Binary
Consistent	Pro-social	159	13.0	13.0	12.7
	Pro-self	94	9.4	9.3	9.0
Inconsistent	Pro-social	46	9.7	10.1	11.3
	Pro-self	43	9.4	10.3	11.1

We turn our attention now to conditional contributions. In Figure 3.2 we plot the conditional contribution of pro-social and pro-self subjects in each of the three treatments. As you can see there is no evidence of any systematic impact of SVO information on contributions. This is confirmed by pairwise treatment comparisons for each possible leader contribution ($p > 0.1$, Mann-Whitney test). This lack of evidence of a systematic effect of information also extends to controlling for inconsistent beliefs. To illustrate, in Table 3.4 we detail mean conditional contribution if the leader contributes 20. There is strong evidence that the conditional contribution of pro-social subjects (if the leader contributes 20) is more than pro-selfs ($p = 0.002$, Mann-Whitney test). This effect, though, is entirely driven by those with a consistent classification. There is no evidence for either those with consistent or inconsistent beliefs that information influenced the conditional contribution ($p > 0.1$, Mann-Whitney tests).

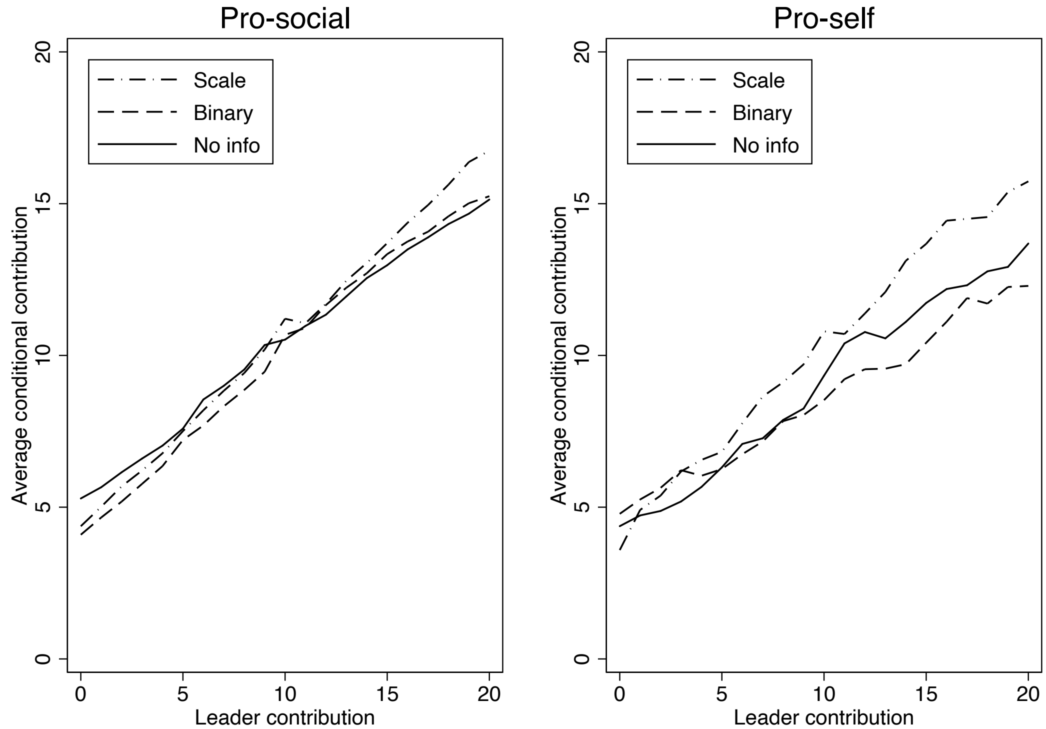


Figure 3.2: Average conditional contribution on leader's contribution by SVO types and treatments

Table 3.4: Mean conditional contribution by SVO belief and classification if the leader contributes 20

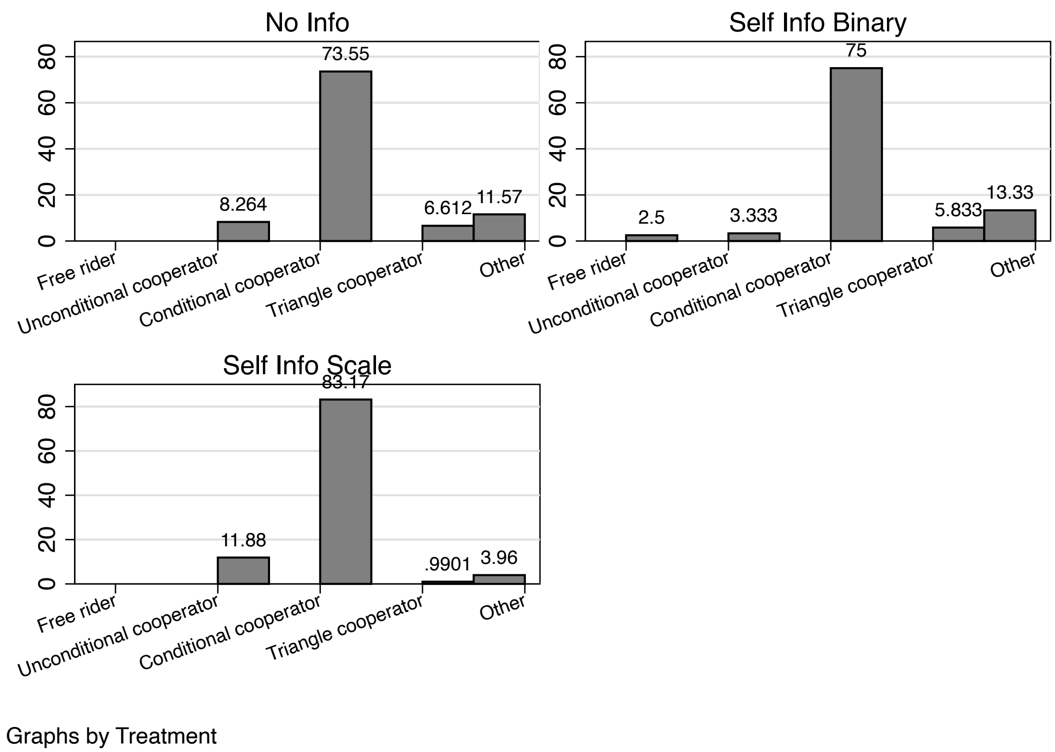
Belief	Classification	No.	Mean contribution		
			No info	Scale	Binary
Consistent	Pro-social	159	15.8	17.6	16.0
	Pro-self	94	12.9	15.6	11.8
Inconsistent	Pro-social	46	12.4	13.2	13.5
	Pro-self	43	15.3	16.4	13.2

We classified subjects based on their entries in the contribution table using the method of Thöni and Volk [2018]. As summarised in Figure 3.3 the large majority of subjects were classified as conditional cooperators. This means their contribution is overall increasing in the contribution of the leader.[5] Around 11% were perfect conditional cooperators meaning they matched the leader contribution. Such high levels of conditional cooperation result in the upward sloping relationships clear in Figure 3.2. We observe very low levels of free-riding, where subjects contribute zero irrespective of the leader contribution.[6] The high level of cooperation we observe is consistent with Hergueux and Jacquemet [2015] who document increased prosociality online. The crucial thing for our purposes is to compare across treatments. As you can clearly see in Figure 3.3 there is no evidence that cooperator type differs across treatments ($p > 0.1$, Proportions test).

[5]For around 57% of these subjects their contribution was monotonically increasing in the leader contribution.

[6]Unconditional cooperators have contributes that are independent of the leader contribution. The average contribution was just under 15. Triangle cooperators have a pattern of increasing and then decreasing contribution as the leader contribution increases. The contribution of those classified as other was decreasing in the contribution of the leader.

Figure 3.3: Distribution of cooperator types by treatment



Result 3.2. *There is no evidence that conditional contributions to the public good are influenced by private information on own SVO.*

We finish the analysis by looking at aggregate contributions across the three treatments. Given that contributions are made independently we simulate aggregate contributions by randomly selecting a leader and three followers and working out aggregate contributions. We then average over 50,000 simulations. The results are summarized in Table 3.5. The table displays average group total contribution from each possible composition of SVO classification within the group. As one would expect, given Results 3.1 and 3.2, there is no evidence of total contributions differing by treatment. The most notable thing is that, ceteris paribus, contributions are around 20% higher if the leader is pro-social rather than pro-self. For instance in a group with 2 pro-social subjects contributions average around 44.6 if the leader is pro-social and 38 if the leader is pro-self. This is driven by pro-social leaders contributing more and then conditional cooperation amplifying this difference.

Result 3.3. *Contributions within a group (containing at least 1 pro-social subject) are, ceteris paribus, higher if the leader is pro-social.*

Table 3.5: Mean conditional contribution by SVO belief and classification if the leader contributes 20

Leader	Followers	Group Total Contribution		
		No Info	Binary	Scale
Pro-social	3 pro-social	47.2	47.2	46.4
Pro-social	2 pro-social	45.9	46.0	45.4
Pro-social	1 pro-social	44.7	44.6	44.5
Pro-social	0 pro-social	43.4	43.4	43.7
Pro-self	3 pro-social	39.5	39.4	38.6
Pro-self	2 pro-social	38.2	38.2	37.8
Pro-self	1 pro-social	36.9	37.2	36.7
Pro-self	0 pro-social	35.6	36.0	35.9

3.6 Conclusion

In this experiment we have looked at the effect of private information on contributions to a public good. We argued that private information should have no effect on contributions if individual's behaviour is driven by belief based preferences, but should have an effect if behaviour is driven by internalized social norms. We report the results of an experiment in which we compare three treatments - one with no information on SVO, one with a binary pro-social or pro-self classification, and one with SVO measured on a scale from very pro-self to very pro-social. We find that contributions are similar across treatments and show no signs of being influenced by private information on SVO. We interpret this as evidence in support of psychological game theoretic models of belief based preferences, such as guilt aversion and reciprocity.

In application our results would suggest that behaviour change is best facilitated by changing first and second order beliefs. In other-words there needs to be a collective shift in what people expect others to contribute, and what they believe others expect them to contribute [Dufwenberg et al., 2011]. That basic notion, of itself, is not particularly novel [Bicchieri, 2016] but there is subtle implication of our findings. Interventions that act at the level of the individual and change an individual's perception of 'the right thing for her to do in this situation' may fail because they act on internalized norms. Interventions aimed at a group level that change, and are seen to change, group perceptions are likely more effective because they can collectively shift beliefs [Ellingsen et al., 2012]. Group interventions may, thus, be more effective than individual interventions.

A number of arguments could be put forward to challenge our interpretation of the evidence and offer alternative explanations for why we observe no effect of private information. In the remainder of this concluding discussion we go through these arguments and set out why we feel they are not compelling. We begin with the possibility that our experiment does not have enough power to detect an effect of private information. This argument seems easy to address given that we have a relatively large number of observations (a total of 342 independent observations with a minimum of 34 per cell) and observe no signs of any effect at all. The only minor caveat is that we see some suggestion of an effect of private information on contributions for subjects whose beliefs are inconsistent with SVO classification. The effect we observe, however, is in the opposite direction to that predicted by the internalized norm approach and so likely reflects some reaction to inconsistent information.

A second potential concern is that subjects misunderstood the SVO classification or were inattentive when performing the SVO measure and reading the feedback. Again, this concern seems relatively easy to allay. We observed a strong statistical relationship between SVO and contributions to the public good, consistent with previous results [Balliet et al., 2009, Pletzer et al., 2018]. We also see a strong statistical correlation between SVO classification and subjects stated beliefs. Both these things give confidence in the measurement of SVO. Moreover, in a companion study that used a very similar protocol [Cartwright, 2019] we observe a statistically robust and sizable, systematic effect on contributions of *public* information about SVO. We would argue, therefore, that it is difficult to make the case that information on SVO was ignored or missed by participants. The evidence points to private information having no effect.

The preceding paragraphs provide a defense of our experimental design. The final issue we consider is whether the theory, and in particular, Hypothesis 3.2 is 'fair' to the internalized norms approach. Here it is useful to make the distinction between normative and empirical norms or injuctive and descriptive norms [House, 2018]. If the norm is based on an empirical expectation of what others would do in this situation then Hypothesis 3.2 seems straightforward. In particular, we know that pro-social individuals contribute more than pro-self individuals. Telling an individual that they are pro-social or pro-self should, therefore, change the empirical expectation and corresponding norm. To argue otherwise would mean the internalized norm based approach lacks bite to make predictions. If the norm is based on a normative opinion of the right thing to do then there would seem more room to question Hypothesis 3.2. For instance, we might argue that a pro-social individual's normative opinion is not influenced by being told she is pro-social. This, however, is hard to reconcile with the fact that public information does have a large effect on contributions [Cartwright, 2019]. If an internalized normative opinion is not changed by private information on SVO then it should not be changed by that same information being public.

The evidence presented here and in our companion paper [Cartwright, 2019] are fully consistent with a psychological game theory approach built around social preferences. As discussed by Fehr and Schurtenberger [2018] social norms and social preferences are necessarily inter-twined. The basic point underlying the belief based approach is that an individuals emotions and sense of what they 'should do' or 'ought to do' are driven in large part by what they believe others will do or think. For instance, an individuals guilt from not contributing to a public good depends on what she believes others expected her to contribute. This creates a complex inter-dependency between individuals in which actions and beliefs overlay. A fuller understanding of this complex inter-dependency can help inform interventions aimed at promoting pro-sociality.

Chapter 4

Leadership in a public good game: Does it matter if the leader is pro-social or just says they are pro-social?

[0]Work joint with Edward Cartwright (De Montfort University, UK)

[0]Each researchers' contribution: the idea, original experimental design and coding, experiment conducting, analysis by Yidan; theory, model suggested and help by Prof. Cartwright.

[0]Funding for this project was provided by Department of Economics and Marketing, De Montfort University, Leicester, UK. & PhD Research Allocation, School of Economics, University of Kent, UK.

4.1 Introduction

The effect of leadership has been studied widely in sequential public goods game. Several studies show in experiments that the presence of leadership can have an effect to boost overall contribution in groups [e.g. Levati et al., 2007, Moxnes and Van der Heijden, 2003, Bahbouhi and Moussa, 2019]. The efficiency enhancing effect of leadership depends on a number of factors. For instance, there is evidence showing a positive effect on group behaviour if leaders have a power to exclude group members Güth et al. [2007] (while Levati et al. [2007] found no effect regarding to this aspect). Of particular interest to us here is the pro-sociality of the leader. Gächter et al. [2012] noticed in their experiment that groups gain higher contribution levels when the leader is cooperatively inclined. Endogenous leadership also outperforms randomly appointed leadership [Haigner and Wakolbinger, 2010, Rivas and Sutter, 2011] (though Arbak and Villeval [2013] found out that it is not necessarily better). This could be because endogenous leaders are more pro-social [Cartwright et al., 2013].

The evidence clearly suggests that the effect of leadership on group cooperation can be positive or negative depending on the leader. The 'right kind of leader' appears to be someone who is pro-social and willing to contribute a large amount to set a good example [Cartwright et al., 2013]. The 'wrong type of leader' is pro-self and sets a bad example. In this chapter we explore what happens if leaders are able to communicate they are pro-social to the group. Does the mere fact a leader says they are pro-social result in a collective shift in beliefs independent of the contribution the leader makes? Moreover, does it make a difference if leaders are able to 'lie' and say they are pro-social when they are pro-self? Indeed, do leaders choose to say they are pro-social.

To explore these questions we ran an experiment with three treatments. In all treatments subjects first performed the slider task to determine their own SVO. They were privately informed of SVO being either pro-self or pro-social. We then used the strategy method whereby participants played two roles respectively – leader and follower. A leader was asked how much they wanted to contribute to the public good. Crucially, they also sent a message to follows about their type. The message could be: the leader is pro-social, the leader is pro-self, or the leaders type is not revealed. A follower was asked how much they wanted to contribute for each possible combination of leader contribution and message.

The difference between the three treatments is the potential for 'hiding' or 'lying' about type. In the Truth game the computer randomly determined the message sent which either correctly informed of leader type or said type is not revealed. Thus, there was no chance of deception. In the Hide game the leader chose whether to reveal type or not reveal type. This provides an opportunity for pro-self leaders to strategically hide type from followers. Finally, in the Cheap-talk game leaders could send any message. This provides an opportunity for pro-self leaders to strategically 'lie' and say they are pro-social.

We hypothesized that pro-self leaders would either hide or lie about type. This leads to testable predictions about contributions. For instance, in the case of a pro-social leader we hypothesise that contributions are highest in the Hide game because a pro-social leader, and only a pro-social leader, can signal their type. By contrast, in the case of a pro-self leader we hypothesise contributions are highest in the Cheap-Talk game because the leader can claim to be pro-social. The aggregate effect would then depend on the proportion of pro-social leaders and the relative size of competing factors.

There are several main findings from our analysis. In terms of the choices of messages sent by leaders, we found that the majority of leaders who were pro-social reveal their true type. As predicted, the majority of leaders who were pro-self chose to not reveal their type, they instead chose to reveal as a pro-social (if they were playing the Cheap-talk game) or to hide SVO type.

We found that pro-social followers contribute statistically more than pro-selfs do. As one would expect, the contribution of followers were increasing with the leader contribution. Crucially, we found that the contribution of followers was (for a fixed leader contribution) higher if the leader said that they were pro-social than if the types were unknown, than if the leader said they were pro-self. Thus, the message sent by the leader made a significant difference. Additionally, if leader type is unknown, we found that pro-selfs treated the leader 'as if' pro-self while pro-socials treated the leader 'as if' pro-social.

When we focus on aggregate contributions we get a complex picture. We found that a greater number of pro-social followers in group increases contributions. Also the leader saying they are pro-social increases group contributions. Aggregated contributions are higher in the Truth game than others if the leader is pro-social. By contrast, aggregated contributions are highest in the Hide than Cheap-talk than Truth game if the leader is pro-self. Ultimately, aggregated contributions are highest overall in the Cheap-talk game. This is primarily because we observe a high level of leaders saying they are pro-social (over 70%) in this treatment. Note that this level is above the proportion of subjects that were pro-social (around 65%) and so is inevitably higher than the proportion of pro-social messages in the Truth and Hide games. Followers do react less to a pro-social message in the Cheap-talk game than the other games. The effect, however, is not enough to offset the higher proportion of pro-social messages.

The chapter is arranged in the following ways: Section 4.2 firstly describes the theoretical model used; Section 4.3 explains several predictions that we have according to our model; Section 4.4 gives the experimental design of present study, as well as the setting and procedure of the experiment; Section 4.5 explains in turn the results; Section 4.6 have a discussion on all results from present study.

4.2 Game and notation

There exists a group of $N = \{1, ..., n\}$ players. Prior to the game each player is tested for social value orientation (SVO) and classified as either pro-social or pro-self. Let t_i denote the type of player $i \in N$ where $t_i = 1$ if player i is pro-social and $t_i = 0$ if player i is pro-self. The players are informed of their own type and then take part in a public good game. Note that players are not aware that SVO could potentially be revealed before the SVO test takes place. We, therefore, view the SVO test as non strategic and not part of the game.

Each player has endowment E and can contribute to a public good. Player 1 is the leader and makes her contribution before the other $n-1$ players. Let $a_1 \in [0, E]$ denote the leader's contribution. The leader also sends a message to the follower that can have the following form: (Sc) the leader is pro-social, (Sf) the leader is pro-self, or (Nr) the leader's type will not be revealed. Let $r_1 \in (Sc, Sf, Nr)$ denote the message sent.

Each of the followers, players $2, .., n$, are told the contribution of the leader, a_1, and the message sent by the leader r_1. This information is common knowledge. Subsequently players $i > 1$ simultaneously and independently make their contributions to the public good. Let $a_i(a_1|r_1) \in [0, E]$ denote the conditional contribution of player i. The contribution of player i is conditional on the leader having contributed a_1 and sent message r_1.

If the leader contributes a_1 and sends message r_1 then total contributions to the public good are

$$A(a_1|r_1) = a_1 + \sum_{i=2}^{n} a_i(a_1|r_1). \tag{4.1}$$

The final monetary payoff of the leader is given by,

$$\pi_1 = E - a_1 + mA(a_1|r_1) \tag{4.2}$$

and that of follower $i > 1$ by,

$$\pi_i = E - a_i(a_1|r_1) + mA(a_1|r_1) \tag{4.3}$$

where $0 < m < 1$ is the marginal per-capita return from the public good.

We compare three different settings in terms of the message that leaders can send. These can be summarised:

- *Truth game*: A third-party (i.e. the computer in our experiment) randomly determines whether the type of the leader will be revealed. It is revealed with probability q (independent of the type of the leader). If type is not revealed then message Nr is sent to followers. If type is revealed and the leader is pro-social ($t_1 = 1$) then message Sc is sent to followers. If type is revealed and the leader is pro-self then message ($t_1 = 0$) is sent.

- *Hide game*: If the leader is pro-social ($t_1 = 1$) then she chooses whether to send the message Sc or Nr. If the leader is pro-self ($t_1 = 0$) then she chooses whether to send the message Sf or Nr.

- *Cheap-talk game*: The leader, irrespective of her type chooses whether to send message Sc, Sf, or Nr.

Note that in the Truth game and the Hide game the message is credible in the sense that only a pro-social leader can send a pro-social message and only a pro-self leader can send a pro-self message. The difference between these two treatments lies in whether the decision to not reveal choice is made randomly (Truth game) or by the leader (Hide game). In the Cheap-talk game leaders can 'lie' by revealing a different type.

4.3 Hypotheses

We focus first on followers. Suppose the leader has contributed a_1 and sent message r_1. Let $b_{ij}(a_1, r_1) \in [0, E]$ denote player i's first order belief on the contribution of player $j > 2, j \neq i$. This belief is conditional on the leader having contributed a_i and sent message r_1. let

$$B_i(a_1, r_1) = \frac{1}{n-2} \sum_{j=2, j \neq i}^{n} b_{ij}(a_1, r_1) \tag{4.4}$$

denote player i's average first order belief. In interpretation, $B_i(a_1, r_1)$ is the average amount she expects other followers to contribute. Let $c_{iji}(a_1|r_1) \in [0, E]$ denote player i's second order belief on $b_{ji}(a_1|r_1)$ for any $j \neq i$. Finally, let

$$C_i(a_1|r_1) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} c_{iji}(a_1|r_1) \tag{4.5}$$

denote player i's average second order belief. In interpretation, C_i is the contribution that she believes others (on average) expect her to make.

There is extensive experimental evidence that contributions in public good games are increasing in first and second order beliefs [Dufwenberg and Kirchsteiger, 2004, Fischbacher and Gächter, 2010, Dufwenberg et al., 2011, Smith, 2013] [Add Fischbacher and Gachter 2010, Dufwenberg et al 2011, Smith 2013]. To put forward testable hypothesis we consider what effect the contribution and message of the leader may have on beliefs. The effect of the leader contribution would seem relatively straightforward.

Assumption 4.1. *$B_i(a_1, r_1)$ and $C_i(a_1, r_1)$ are both a weakly increasing function of a_1 for all $i > 1$ and any r_1.*

Assumption 1 leads directly to our first hypothesis.

Hypothesis 4.1. *The contribution of followers are increasing in the leader contribution a_1 for any r_1.*

The existing evidence would give strong support to Assumption 4.1 and Hypothesis 4.1.

Our primary focus will be on the message sent by the leader and type of game. We begin with the conjecture that the type of leader creates a frame of reference that can influence beliefs. To explore the implications of this let p_i denote the probability that player $i \in N$ puts on the leader being pro-social. (Meaning $1 - p_i$ is the probability the leader is pro-self.) Then, let $B_i(a_1, G|p_i)$ and $C_i(a_1, G|p_i)$ denote the first and second order beliefs of player i conditional on the leader contributing a_1, sending message G, and i believing the leader is pro-social with probability p_i. A key assumption we make is that first and second order beliefs are increasing in p_i.

In other words, ceteris paribus, a player expects higher contributions if the leader is pro-social.

Assumption 4.2. *$B_i(a_1, G|p_i)$ and $C_i(a_1, G|p_i)$ are both a weakly increasing function of p_i for all $a_1 \in [0, E], G \in \{Sf, Sr, Nr\}$ and $i > 1$.*

This assumption leads directly to a key hypothesis.

Hypothesis 4.2. *The contribution of followers is, for a fixed leader contribution a_1, (a) higher if the leader can be inferred to be pro-social than if leader type is unknown, and (b) higher if type is unknown than if the leader can be inferred to be pro-self.*

To progress further we need to consider the incentives of the leader. Leader contribution may be influenced by her first and second order beliefs. It might also be influenced by strategic concerns and setting a good example [Cartwright and Patel, 2010]. Key for us, though, is to think through the consequences of Assumption 4.2. The following hypotheses follow directly.

Hypothesis 4.3. *Leaders who are pro-social will want to reveal their type where possible.*

Hypothesis 4.4. *Leaders who are pro-self will want to not reveal their type where possible.*

In applying Hypotheses 4.3 and 4.4 consider Table 4.1. Here we detail intuitive assumptions on p_i, the probability the leader is inferred to be pro-social by player i, given the message sent by the leader and the type of game. In the Truth game an Sc message reveals the leader is pro-social and so we naturally expect $p_i = 1$. Similarly an Sf message reveals the leader is pro-self and so we expect $p_i = 0$. An

Nr message conveys no information about leader type and so we set $p_i = \overline{p}_i$ where $\overline{p}_i$ is player i's prior on the proportion of people in the population who are pro-social.

In the Hide game an Sc message again reveals the leader is pro-social and so we naturally expect $p_i = 1$. Similarly an Sf message reveals the leader is pro-self and so we expect $p_i = 0$. In this game, however, applying Hypotheses 4.3 and 4.4, we can infer that a type Nr message is likely to imply the leader is pro-self. Hence we set $p_i = 0$. Finally, consider the Cheap-talk game. Here, Hypotheses 4.3 and 4.4 would imply all leaders send message Sc. Thus, a message of Sc conveys no information and we set $p_i = \overline{p}_i$. Table 4.1 sets out important predictions across the three treatments which lead to our next Hypothesis.

Hypothesis 4.5. *Followers will: (a) Contribute more in the Truth and Hide game than the Cheap-talk game if the leader says they are pro-social (for a fixed leader contribution a_1). (b) Contribute more in the Truth game than the Hide or Cheap-talk game if the leader type is not revealed (for a fixed leader contribution a_1).*

Table 4.1: Predicted probability on the type of the leader conditional on the message sent by the leader and the type of game.

Message	p_i		
	Truth	Hide	Cheap-talk
Sc	1	1	$\overline{p}_i$
Nr	$\overline{p}_i$	0	0
Sf	0	0	0

In formulating a hypothesis on aggregate contributions we note two competing forces. On the one hand, as we can see in Table 4.1, the Cheap-talk game provides no way for a pro-social leader to signal type. If, therefore, there is a big gap between follower contributions if the leader is known to be pro-social ($p_i = 1$) versus of unknown type (say, $\overline{p}_i = 0.5$), then contributions would be predicted to be lowest in the Cheap-talk game. On the other hand, the cheap-talk game provides a way for pro-selfs to 'disguise' their type. If, therefore, there is big gap in follower contributions if the leader is unknown type ($\overline{p}_i = 0.5$) versus known to be pro-self ($p = 0$) then contributions would be highest in the Cheap-talk game. The overall effect is, thus, ambiguous. We can, though, make two final hypotheses that capture the preceding logic. Our next hypothesis recognises that a pro-social leader has more opportunity to signal type in the Hide than Truth than Cheap-talk game.

Hypothesis 4.6. *If the leader is pro-social then aggregate contributions will be higher in the Hide than Truth and Cheap-talk game.*

Our final hypothesis recognises that a pro-self leader has more opportunity to 'disguise' type in the Cheap-talk than Truth than Hide game

Hypothesis 4.7. *If the leader is pro-self then aggregate contributions will be higher in the Cheap-talk than Truth than Hide game.*

4.4 Experimental Design

At the start of the experiment we determined the SVO of subjects using the slider method of Murphy et al. [2011]. This method consists of 6 questions in which subjects have to determine an allocation of money between themselves and someone else. A slider angle is determined from the outcome of the 6 questions. Those with a slider angle below 22.45° are classified as pro-self and those with an angle above (including) 22.45° as pro-social. Before subjects were told the outcome of the SVO test they were asked what type they thought they would be. This allows us to control for 'surprise' at SVO classification.

In the public good game we set $n = 4$ and $E = 5$. We used the strategy method adapted from Fischbacher et al. [2001] in which subjects are told that the leader will be randomly chosen after choices have been elicited for all possible outcomes. Hence, subjects first state an unconditional contribution and clarify what message they would send if they are the leader - in the Truth game the message of revealing or not is determined with probability $q = 0.7$. Subjects then fill in a contribution table that states the amount they would contribute for any contribution of the leader $a_1 \in \{0, 1, .., 5\}$. Moreover, they fill in three contribution tables, corresponding to each possible message the leader may send: Sc, Sf, or Nr. The order of possible messages and contributions were randomly shuffled by the program for each subject to avoid order effects.[1]

Subjects were recruited from Prolific[2] and the experiment was run using O-Tree [Chen et al., 2016]. We selected subjects who were resident in the UK and are currently in employment. Each subject goes through the experiment at their own speed with the matching to groups and determination of leader performed afterwards. Subjects received a fixed payment of £6.85/hr for participation and a bonus payment based on the outcome of the public good game. The experiment took around 7 minutes for participants to finish.

Our experiment was run in two stages. The first stage was run in March 2021, where 125 subjects took part in the Cheap-talk treatment, and 125 in the Hide treatment. The second stage was run in August 2021, with 82 subjects in the Cheap-talk, 81 in the Hide and 200 in the Truth treatments. In total, we have 207 subjects in the Cheap-talk, 206 in the Hide and 200 in the Truth treatments. Overall, 66% of subjects were female, 45% under 30, 61% with income above

[1] See Appendix for screenshots.

[2] https://www.prolific.co

£30,000 and 61% with a university degree.

4.5 Results

4.5.1 Descriptive Statistics

Table 4.2 provides a summary of descriptive statistics across the three treatments. We can observe that the majority of participants are classified as pro-social both from the SVO test (around 65% of subjects) and self-valuation (around 62%). The average SVO angle is around 25.2°so above the 22.45°threshold, indicating that an average participant in the experiment is pro-social. The three treatments have a similar distribution of slider and angle, levels of pro-sociality and demographics according to a one-way ANOVA. The one exception is gender where we observe a lower proportion of females in the Truth treatment. We will control for this in the regression analysis.

Table 4.2: Descriptive statistics

	Cheap-talk	Hide	Truth	p-value
Average slider angle	25.66	25.61	24.77	*0.74*
Pro-socials (%)	68.12	67.48	63.50	*0.57*
Self-valued Pro-socials (%)	65.22	60.68	60.00	*0.50*
Female (%)	68.12	73.79	56.00	*0.004*
Age 18 $\sim$ 40 (%)	76.33	72.33	66.50	*0.10*
Employed (full-time) (%)	72.46	64.08	65.50	*0.95*
Undergraduate or above (%)	39.13	44.66	35.00	*0.42*
Income above £30,000 (%)	60.87	59.23	62.50	*0.68*
Observations	207	206	200	

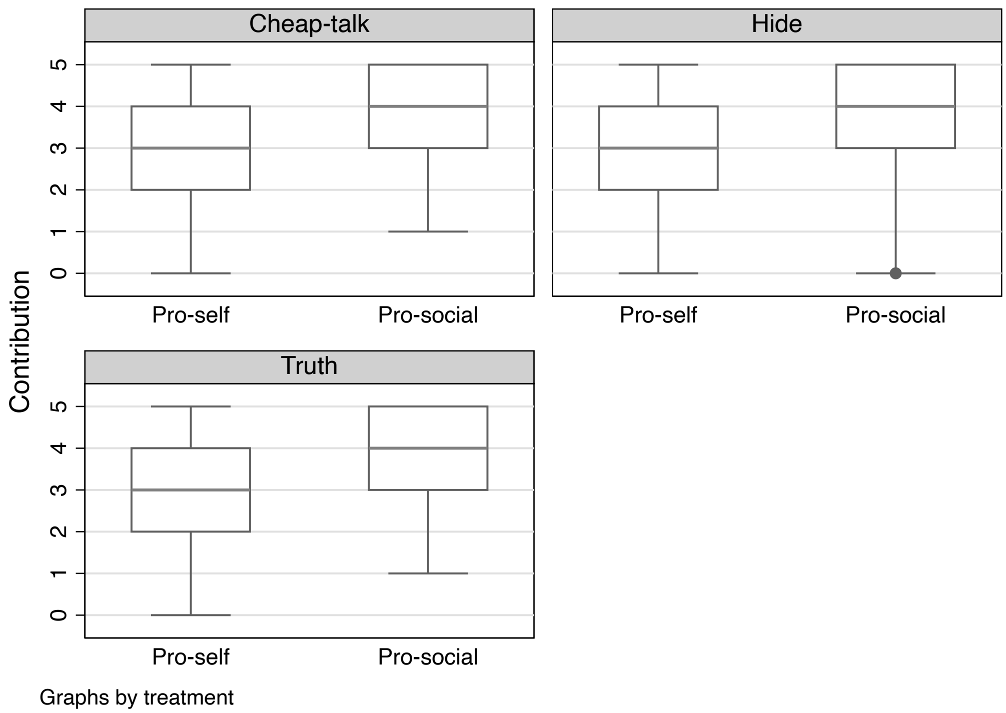
The last column shows the p-values of oneway ANOVA Bartlett's test for equal variances.

4.5.2 Leader choices

Next, we analyze the leader, unconditional contribution in the public goods game. Figure 4.1 displays the distribution of leaders' contributions by the two SVO types. As the plot shows, there is an apparent difference between these two groups in terms of their contribution. Pro-socials exhibit a higher level of contribution compared to those classified pro-selfs. A two-sample Wilcoxon rank-sum (Mann-Whitney) test suggests a statistically significant difference between types at the 5%

level for all treatments. The finding is in line with the results found in our previous Chapters 2 and 3.

Figure 4.1: Contribution by SVO



In addition to the contribution decision, leaders were also asked to choose whether and how would reveal their SVO type to followers in the group. Recall, the message sent can be represented as 'Sc', 'Sf' and 'Nr', corresponding to message the leader is pro-social, the leader is pro-self, or the leader's type is not revealed. A summary of the message sent as well as average contribution conditional on message sent is provided in Table 4.3. The data in Table 4.3 allows us to evaluate Hypotheses 4.3 and 4.4 which said, respectively, that pro-socials will reveal type while pro-selfs will hide type.

Table 4.3: Leader messages

	Total	Messages		
		Cheap-talk		
		'Sc'	*'Sf'*	*'Nr'*
Pro-social #	141	113	1	27
		3.88 (.10)	4.00	3.93 (.18)
Pro-self #	66	34	8	24
		2.97 (.28)	3.13 (.40)	2.88 (.23)
		Hide		
		'Sc'	*'Sf'*	*'Nr'*
Pro-social #	139	104	-	35
		4.08 (.10)	-	3.31 (.23)
Pro-self #	67	-	20	47
		-	3.65 (.28)	2.66 (.19)
		Truth		
		'Sc'	*'Sf'*	*'Nr'*
Pro-social #	127	87	-	40
		3.94 (.10)	-	4.10 (.14)
Pro-self #	73	-	53	20
		-	2.89 (.18)	3.45 (.29)

In the Cheap-talk treatment we see that 80.14% of pro-socials and 52.52%

of pro-selfs send message ‘Sc’ compared to only 1% and 12%, respectively, that send message ‘Sf’. This is in line with Hypotheses 4.3 and 4.4. There is evidence, however, that a significant proportion of participants choosing to not reveal their type. This could be a desire to not reveal information or, in the case of pro-selfs, to not lie. Those that do reveal type, however, overwhelmingly send message ‘Sc’. A similar picture emerges in the Hide treatment. Specifically, 75% of pro-socials choose to reveal type compared to only 30% of pro-selfs. We, thus, find strong support for Hypotheses 4.3 and 4.4.

We can look next at mean leader contribution as a function of message sent. In the Cheap-talk game there is no evidence that pro-social leaders or pro-self leaders who do not reveal their type contribute any more or less than those who reveal type. In the Hide game, by contrast, there is evidence that those who do not reveal type contribute less than those who reveal type (Mann-Whitney test $p = 0.0001$). Of particular interest is whether pro-selfs who sent the message ‘Sc’ or ‘Nr’ contributed more than those who revealed ‘Sf’. You can clearly see in Table 4.3 that there is no evidence of this at all. Indeed, pro-selfs who ‘honestly’ reveal type contribute *more* than those who hide or lie about type in both the Hide and Cheap-talk games. Claiming to be pro-social does not, therefore, appear to change the behaviour of the leader. This shows up most clearly in the Cheap-talk game and the fact that pro-social leaders who send message ‘Sc’ contribute significantly more than pro-self leaders who send message ‘Sc’ (Mann-Whitney test, $p < 0.05$).

We can summarise with our first two results:

Result 4.1. *The majority of leaders who are pro-social reveal their type.*

Result 4.2. *The majority of leaders who are pro-self do not reveal their type. Pro-self leaders who hide or lie about type contribute less than pro-social leaders.*

4.5.3 Followers

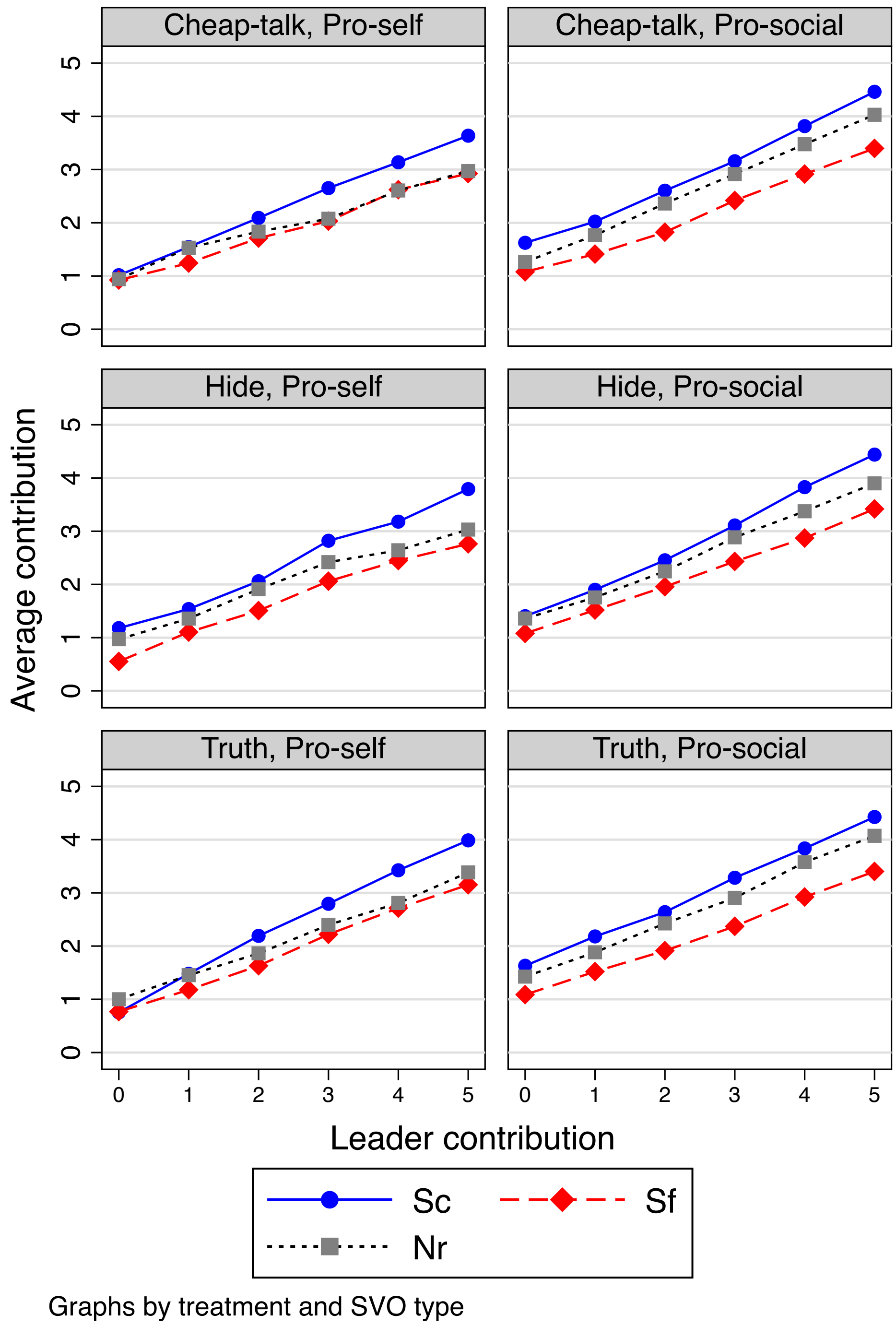
This section discusses the conditional contribution by followers. Recall that during the contribution stage, participants were asked, as followers, to input conditional contributions given each possible leader's message and contribution. The order of possible types and contribution levels were randomly shuffled for each participant to avoid order effects.

4.5.3.1 Conditional Contributions

Figure 4.2 plots the mean conditional contribution by followers as a function of the leader's contribution and the message sent by the leader. We have sub-plots for each of the three treatments and type of the follower. There are several patterns can be observed directly in the figure. First, it is directly apparent that the mean conditional contribution is increasing with the contribution of leader for both pro-selfs and pro-socials irrespective of the message or treatment. This supports our Hypothesis 4.1.

Result 4.3. *The contribution of followers are increasing in the leader contribution.*

Figure 4.2: Conditional contribution of followers as a function of leader contribution, split by follower SVO type and treatment



Another result apparent in Figure 4.2, if we compare the right hand sub-plots with the left-hand subplots, are that the contribution levels of pro-social followers are generally greater than those of pro-selfs, everything else the same. This suggests that pro-socials are more reciprocal than pro-selfs, echoing findings in Van Lange [1999]. Mann-Whitney test results are summarized in Table 4.4 where the conditional contributions by these two types are compared and tested for a fixed leader contribution. As the table shows, the differences in conditional contributions by the two types are highly statistically significant.

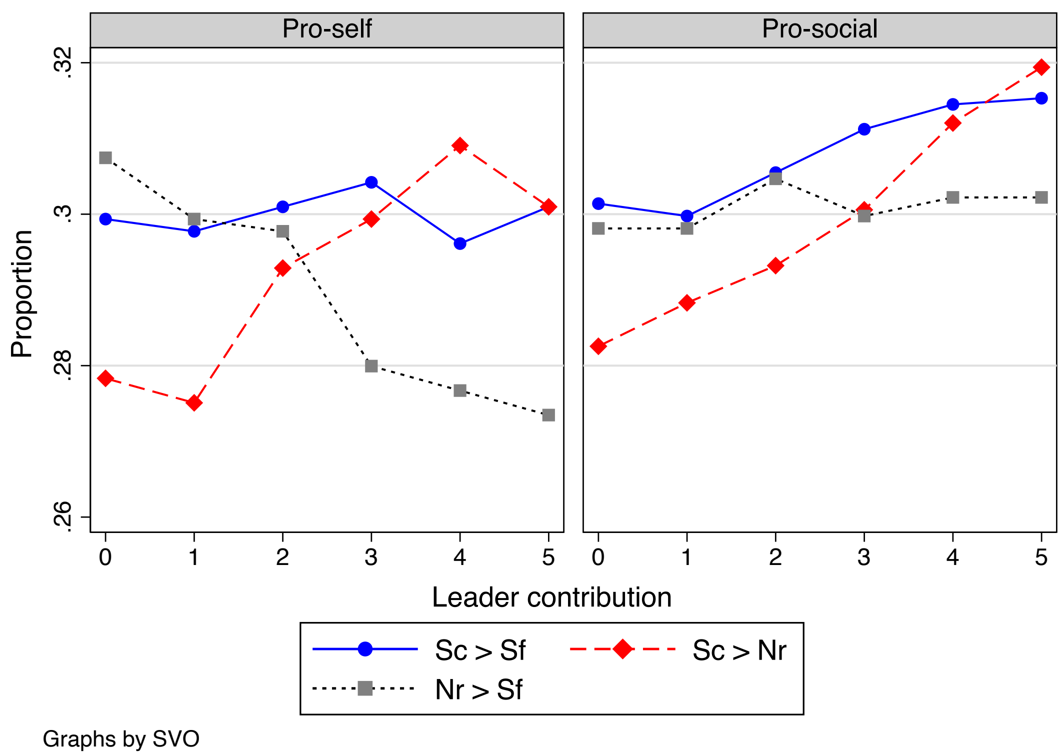
Table 4.4: Mann-Whitney tests between SVO types on conditional contribution as a function of leader contribution

Mann-Whitney Test Results						
Leader contribution	Sc		Sf		Nr	
	Z	p-value	Z	p-value	Z	p-value
0	-3.946	*0.0001*	-2.601	*0.0093*	-2.883	*0.0039*
1	-4.052	*0.0001*	-2.521	*0.0117*	-3.136	*0.0017*
2	-4.385	*0.0000*	-2.914	*0.0036*	-4.876	*0.0000*
3	-4.588	*0.0000*	-2.645	*0.0082*	-6.238	*0.0000*
4	-5.980	*0.0000*	-2.457	*0.0140*	-7.089	*0.0000*
5	-5.438	*0.0000*	-3.176	*0.0015*	-6.580	*0.0000*

Result 4.4. *Pro-social followers, ceteris paribus, contribute significantly more than pro-selfs irrespective of the leader contribution or leader message.*

A final thing to consider in Figure 4.2 is the relationship between conditional contribution and leader message. We can see that the average contribution level is the highest for both SVO types if they were told that the leader is pro-social, whereas it is the lowest if they were told the leader is pro-self. This is consistent with Hypothesis 4.2. We note, however, that this difference in means is not reflected across all subjects. To illustrate, Figure 4.3 shows the proportion of subjects who contributed more in three scenarios: (1) more if they were told that their leader is pro-social than if leader is pro-self; (2) more if leader is pro-social than if leader is unknown; (3) more if leader is pro-self than if leader is unknown. We can see in all cases that it is around 30% of subjects condition their contribution on the message sent. This means that in the majority of instances the contribution is not dependent on the message.

Figure 4.3: Proportion of subjects contributed more in three scenarios



Further tests were conducted to test for a significant difference in contribution between each pair of messages sent for the two SVO types in each treatments. The results suggests that subjects contributed significantly more if the message is 'Sc' rather than 'Sf' ($p < 0.01$). Pro-social leaders also contribute more when the leader sends message 'Nr' rather than 'Sf' ($p < 0.05$). One interesting thing to observe in both Figures 4.2 and 4.3 is the way in which subjects respond to an 'Nr' message. There is a suggestion that pro-selfs react in a similar way to an 'Nr' message as they do an 'Sf' message and that pro-socials react in a similar way to an 'Nr' message as they do an 'Sc' message. This suggests that subjects 'subconsciously' think that other person is of the same type as them and act upon this assumption. This echos the results of previous studies [Kelley and Stahelski, 1970, Liebrand, 1984, Van Lange, 1992, Smeesters et al., 2003].

Result 4.5. *The contribution of followers is (for a fixed leader contribution) higher if the leader sends message 'Sc' than 'Nr' than 'Sf'.*

Result 4.6. *Pro-selfs tend to react to message 'Nr' in the same way as message 'Sf' while pro-socials tend to react to message 'Nr' in the same way as message 'Sc'.*

We now turn to a test of Hypothesis 4.5. First, we consider whether contributions are higher in the Truth and Hide game than the Cheap-talk game if the leader says they are pro-social. In Table 4.5 we provide the results of a linear regression with Bootstrap replications and demographics as extra controls.[3] The dependent variable is contribution conditional on message 'Sc'. The results show that, when fixing leader's contribution, followers generally contribute more in the Truth game if

[3]Bootstrap is a statistical method of expanding samples, which provides a good idea for solving the problem of small samples in our case. In analysis, 100 replications were conducted with other settings default.

receiving ‘Sc’ message, though this difference is statistically insignificant. Moreover, the Hide and Cheap-talk games have very close contribution levels. We, thus, fail to find support for Hypothesis 4.5(a).

Table 4.5: Regression on followers' contribution (Sc)

	(1)		(2)	
	Average Follower contribution		Margins	
Hide	-0.0827	(-0.48)	2.721***	(73.09)
Truth	-0.0994	(-0.61)	2.802***	(73.65)
Leader Con=1	0.440***	(2.80)	1.858***	(30.83)
Leader Con=2	1.010***	(6.53)	2.413***	(50.34)
Leader Con=3	1.565***	(11.39)	3.038***	(80.17)
Leader Con=4	2.169***	(15.08)	3.633***	(79.62)
Leader Con=5	2.768***	(18.77)	4.230***	(81.84)
Hide × Leader Con=1	0.0118	(0.06)		
Hide × Leader Con=2	-0.0145	(-0.07)		
Hide × Leader Con=3	0.119	(0.62)		
Hide × Leader Con=4	0.117	(0.58)		
Hide × Leader Con=5	0.130	(0.64)		
Truth × Leader Con=1	0.175	(0.85)		
Truth × Leader Con=2	0.155	(0.75)		
Truth × Leader Con=3	0.230	(1.21)		
Truth × Leader Con=4	0.206	(1.02)		
Truth × Leader Con=5	0.187	(0.86)		
Gender	-0.0688	(-1.61)		
Age range	0.0305*	(1.84)		
Employment status	-0.0112	(-0.98)		
Income range	-0.0201	(-1.06)		
Education level	-0.0458***	(-2.89)		
SVO type belief	0.492***	(9.08)		
Cheap-talk			2.743***	(81.89)
Leader Con=0			1.357***	(17.10)
Constant	0.873***	(5.35)		
Observations	3678		3678	

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Next we consider whether followers' contribution is greater in the Truth game than other two games if they received an 'Nr' message. The results of a regression which follower contribution conditional on message 'Nr' are summarized in Table 4.6. Again, we find no statistically significant difference across treatments. We, thus, fail to find support for Hypothesis 4.5(b). Overall, while contributions are in the direction posited by Hypothesis 4.5 the differences are very small. It seems, therefore, that followers are more reactive to the message and not the treatment (which may inform on the reliability of the message).

Table 4.6: Regression on followers' contribution (Nr)

	(1)		(2)	
	Average Follower contribution		Margins	
Hide	0.0877	(0.56)	2.416***	(65.58)
Truth	0.130	(0.84)	2.517***	(69.56)
Leader Con=1	0.531***	(3.51)	1.680***	(32.25)
Leader Con=2	1.034***	(7.30)	2.183***	(46.58)
Leader Con=3	1.488***	(11.34)	2.700***	(60.11)
Leader Con=4	2.039***	(13.29)	3.209***	(56.93)
Leader Con=5	2.531***	(17.35)	3.708***	(62.12)
Hide × Leader Con=1	-0.138	(-0.69)		
Hide × Leader Con=2	-0.131	(-0.73)		
Hide × Leader Con=3	0.0121	(0.07)		
Hide × Leader Con=4	-0.136	(-0.63)		
Hide × Leader Con=5	-0.148	(-0.70)		
Truth × Leader Con=1	-0.0764	(-0.38)		
Truth × Leader Con=2	-0.0838	(-0.43)		
Truth × Leader Con=3	-0.0379	(-0.21)		
Truth × Leader Con=4	-0.0136	(-0.07)		
Truth × Leader Con=5	0.0186	(0.09)		
Gender	-0.0694*	(-1.66)		
Age range	0.0621***	(3.65)		
Employment status	-0.0138	(-1.22)		
Income range	-0.0297	(-1.52)		
Education level	-0.0258*	(-1.67)		
SVO type belief	0.574***	(13.94)		
Cheap-talk			2.419***	(60.86)
Leader Con=0			1.220***	(17.56)
Constant	0.348*	(1.96)		
Observations	3678		3678	

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 4.7. *We find no significant difference in followers contribution across the Truth, Hide and Cheap-talk games conditional on the leader sending an 'Sc' or 'Nr' message.*

4.5.3.2 Cooperator Types

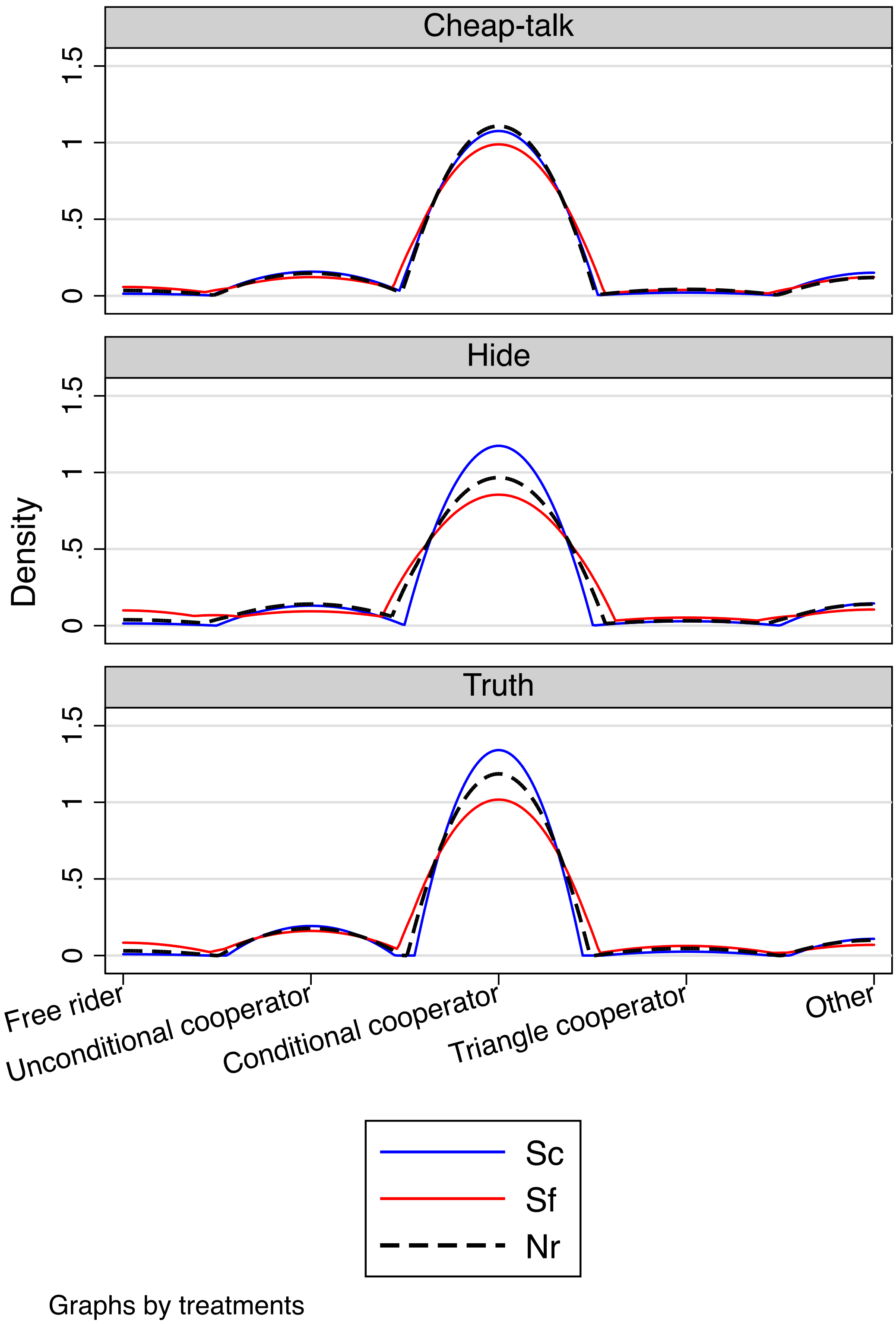
In this section, we investigate cooperator types on the basis of conditional contributions. We first classify participants into a cooperator type according to their contribution as a function of the leader's conditions. Recall that there are three tables for followers to indicate their contribution, each form is designed for one of leader's messages: 'Sc', 'Sf' or 'Nr'. The results of the classification are displayed in the following Table 4.7.

Table 4.7: Conditional cooperator types by leader types

Conditional Cooperator Type	Frequency		
	Sc	Sf	Nr
Free rider	5 (0.82%)	38 (6.20%)	15 (2.45%)
Unconditional cooperator	64 (10.44%)	58 (9.46%)	66 (10.77%)
Conditional cooperator	479 (78.14%)	446 (72.76%)	463 (75.53%)
Triangle cooperator	10 (1.63%)	24 (3.92%)	17 (2.77%)
Other	55 (8.97%)	47 (7.67%)	52 (8.48%)

As the table shows, most of the participants, on average 75%, are conditional cooperators. This is consistent with findings identified in Chapter 3. Overall, the distribution of cooperator types are generally identical across leader's messages. The one exception is a higher frequency of free-riders when the leader sends 'Sf' and lower frequency if message 'Sc'. The density of cooperator types by treatments and leader's messages is plotted in Figure 4.4. The Y-axis displays the density for each category and the x-axis lists the cooperator type. In the Cheap-talk game we see no discernible difference in types according to message sent. In the Hide and Truth games, by contrast, we see some differences. In particular, the proportion of conditional cooperators is highest with message 'Sc' and lowest with message 'Sf'. These differences in the proportion of free-riders and conditional cooperators are consistent with Result 4.5 and the finding that participants reacted to the message sent by the leader (controlling for leader contribution).

Figure 4.4: Distribution of cooperator types by treatments and leader's messages



Additional logistic regressions were conducted to see if there is a significant difference of the number of conditional cooperators for each leader type across treatments. The results suggest that, with the Cheap-talk game as a baseline, for a fixed message sent, there is no statistically significant treatment effect on whether followers are conditional cooperators or not. The coefficient is $.2421, p = 0.313$ in the Truth game and $.1592, p = 0.498$ in the Hide game when given message 'Sc'. The coefficient is $-.0720, p = 0.749$ in the Truth game and $-.1774, p = 0.422$ in the Hide game given 'Sf'. Finally, $.0365, p = 0.873$ in the Truth game and $-.1608, p = 0.479$ in the Hide game if given 'Nr'. Therefore, we have a result as below.

Result 4.8. *There is no compelling evidence that the message sent by the leader influences the cooperator type of the follower.*

4.5.4 Group efficiency

In this section we look at overall contributions to the public good. We simulated the group to generate aggregated total contributions across different scenarios by treatment. The simulation was conducted by Python, using a program that randomly selected four participants as a group and one of them as the leader in the group. The group total contribution then consisted of the unconditional contribution from leader and the corresponding conditional contribution by followers given the message sent by the leader. During the simulations, the number of pro-social followers in group and the message sent by the leader is also noted by the script.

Theoretically we have more than millions group combinations for each treatment.[4] Therefore, in order to reduce the workload but also making sure that we have enough observations (all group structures appear in the simulation outcome), after several practice, we ran the simulation for around 300,000 times for each treatment.

The outcome predicts, on average, 13.89 of group total contribution in the Cheap-talk game, 11.28 in the Hide game and 12.68 in the Truth game. Generally, the Cheap-talk game has the greatest level of aggregate contribution, while the Hide game has the least. To being to explore this in more detail, Table 4.8 and Figure 4.5 present the outcome of the simulations by number of pro-social followers in the group, treatments and messages sent. In each column of Table 4.8 average group total contribution and standard deviation are displayed by message and the number of pro-social followers in the group. The table is split by treatments.

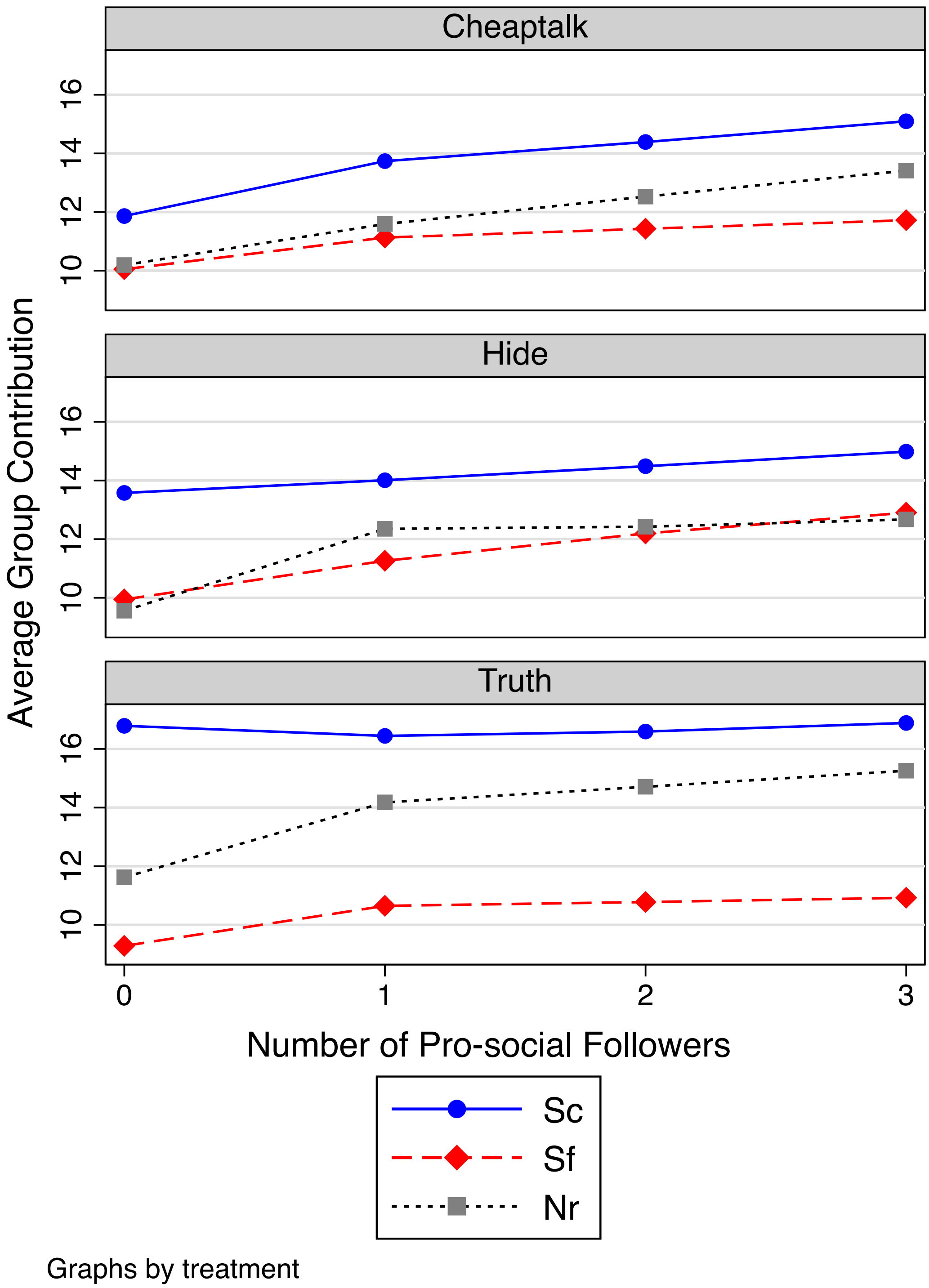
[4]According to Combinatorics, if we want to select 4 samples from 200 objects (order not matters), a quick calculation will give us more than 60 million. While we also need to select one member as a leader, then there will be fourfold possibilities.

Table 4.8: Total group contribution simulation

Number of Pro-social Followers	Cheap-talk			Hide			Truth		
	Sc	Sf	Nr	Sc	Sf	Nr	Sc	Sf	Nr
0	11.86	10.05	10.19	13.58	9.94	9.56	16.79	9.28	11.62
	(3.95)	(3.19)	(3.22)	(1.54)	(3.41)	(3.28)	(2.78)	(3.85)	(4.06)
1	13.74	11.13	11.59	14.01	11.26	12.35	16.45	10.65	14.18
	(3.69)	(3.10)	(3.36)	(3.23)	(2.96)	(2.93)	(3.04)	(3.53)	(3.52)
2	14.39	11.43	12.53	14.49	12.19	12.42	16.59	10.78	14.71
	(4.10)	(3.29)	(3.54)	(3.61)	(3.59)	(3.37)	(3.01)	(3.49)	(3.54)
3	15.09	11.72	13.41	14.98	12.90	12.67	16.89	10.92	15.26
	(4.34)	(3.43)	(3.70)	(3.80)	(3.90)	(3.62)	(2.91)	(3.45)	(3.56)
All	14.62	10.98	12.22	14.65	10.23	10.04	16.68	9.65	13.38
Treatment Average	*13.89*			*11.28*			*12.68*		

It is immediately apparent from Table 4.8 and Figure 4.5 that the total group contribution is the highest if leaders say they are pro-social (14.62, 14.65 and 16.68 on average for Cheap-talk, Hide and Truth game), while the value is generally the lowest if leader say they are pro-self (10.98, 10.23 and 9.65 on average for Cheap-talk, Hide and Truth game). This pattern is reflected in all treatments. As one would expect, given Result 4.3, the group contribution is also increasing in the number of pro-social followers, irrespective of the leader's message or treatment. If there were 3 pro-social followers in the group, the group total contribution on average, after a quick math according to Table 4.8, is 14.74, whereas it is 9.83 when there were three pro-self followers in the group.

Figure 4.5: Average group contribution by treatments and leader messages



If we focus on the situations where leaders say they are pro-socials, the group total contribution is greatest in the Truth game and similar in the Cheap-talk and Hide games. This goes somewhat against Hypothesis 4.5(a). In principle, contributions in the Hide game should be similar to those in the Truth game, given that in both cases an 'Sc' message could only be sent by someone who is pro-social. In practice it seems that followers are more responsive to the 'Sc' message in the Truth game. If we focus on situations where leaders do not reveal type then contributions are, again, greatest in the Truth game and similar in the Cheap-talk and Hide games. This is consistent with Hypothesis 4.5(b). If the leader sends message 'Sf' then contributions are relatively similar across treatments.

In Table 4.9 we present the results of an OLS regression on the simulation data. The margins are also displayed in the table. In the regression, zero number of pro-social follower, Hide and message 'Nr' are the baseline. The estimations confirm that the number of pro-social followers in the group has a significant positive effect on group contribution. Moreover, this influence is the result of aggregation of multiple impacts, which includes the natural tendency of unconditional cooperation by pro-socials, and the behavioural fact that after receiving the message, there is still a greater willingness for pro-socials to cooperate. The results also show that the Truth game has a significantly higher level of cooperation in the group than the Cheap-talk and Hide games. Yet, the effect is a compound impact where we need to split the analysis by messages. The group contribution with message 'Sc' is lower in the Cheap-talk game. In the case of message 'Sf' we find group contribution is not predicted to be lower in the the Cheap-talk game. Finally, for all treatments, there is a positive effect of 'Sc' on group contribution, whereas 'Sf' induces lower contribution.

Table 4.9: Regression on group contribution simulation

	(1)		(2)	
	Total group contribution		Margins	
Pro-social Followers=1	2.791***	(64.82)		
Pro-social Followers=2	2.864***	(155.23)		
Pro-social Followers=3	3.114***	(190.82)		
Cheaptalk	0.633***	(52.88)	12.60***	(2821.77)
Truth	2.065***	(219.44)	13.89***	(5016.49)
Pro-social Followers=1 × Cheaptalk	-1.392***	(-27.77)		
Pro-social Followers=1 × Truth	-0.238***	(-4.88)		
Pro-social Followers=2 × Cheaptalk	-0.530***	(-22.44)		
Pro-social Followers=2 × Truth	0.222***	(9.70)		
Pro-social Followers=3 × Cheaptalk	0.106***	(4.92)		
Pro-social Followers=3 × Truth	0.520***	(24.30)		
Sc	4.018***	(50.32)	15.28***	(1205.58)
Sf	0.385***	(37.22)	10.76***	(1772.35)
Pro-social Followers=1 × Sc	-2.358***	(-25.43)		
Pro-social Followers=1 × Sf	-1.473***	(-13.22)		
Pro-social Followers=2 × Sc	-1.953***	(-23.69)		
Pro-social Followers=2 × Sf	-0.613***	(-13.47)		
Pro-social Followers=3 × Sc	-1.706***	(-20.81)		
Pro-social Followers=3 × Sf	-0.154***	(-3.92)		
Cheaptalk × Sc	-2.346***	(-27.66)	13.64***	(1406.23)
Cheaptalk × Sf	-0.527***	(-21.44)	11.00***	(900.52)
Truth × Sc	1.148***	(13.20)	16.74***	(1353.72)
Truth × Sf	-2.723***	(-198.60)	10.28***	(2061.50)
Pro-social Followers=1 × Cheaptalk × Sc	2.832***	(28.24)		
Pro-social Followers=1 × Cheaptalk × Sf	1.155***	(8.93)		
Pro-social Followers=1 × Truth × Sc	-0.538***	(-5.28)		
Pro-social Followers=1 × Truth × Sf	0.284**	(2.45)		
Pro-social Followers=2 × Cheaptalk × Sc	2.144***	(24.37)		
Pro-social Followers=2 × Cheaptalk × Sf	-0.341***	(-5.97)		
Pro-social Followers=2 × Truth × Sc	-1.329***	(-14.72)		
Pro-social Followers=2 × Truth × Sf	-0.979***	(-20.01)		
Pro-social Followers=3 × Cheaptalk × Sc	1.718***	(19.66)		
Pro-social Followers=3 × Cheaptalk × Sf	-1.394***	(-27.50)		
Pro-social Followers=3 × Truth × Sc	-1.831***	(-20.35)		
Pro-social Followers=3 × Truth × Sf	-1.839***	(-42.36)		
Hide			12.51***	(2965.07)
Nr			12.51***	(3743.88)
Cheaptalk × Nr			11.88***	(2084.61)
Hide × Nr			11.46***	(1505.21)
Hide × Sc			14.27***	(494.70)
Hide × Sf			11.52***	(668.53)
Truth × Nr			13.72***	(2474.64)
Constant	9.558***	(1924.91)		
Observations	5311877		5311877	

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 4.9. *If the leader sends message 'Sc' then aggregate contributions are higher in the Truth than Hide and Cheap-talk games.*

Result 4.10. *If the leader sends message 'Sf' then aggregate contributions are higher in the Hide than Cheap-talk and Truth games.*

It remains to test Hypotheses 4.6 and 4.7. Here we need to take into account the message that the leader will send. Table 4.10 summarizes the aggregated group contribution by leader's SVO type. The last six columns display the group and leader's contribution by messages sent. We include the leader's contribution as a control and generated regression to test the Hypotheses 4.6 and 4.7. The results of the regression as well as the margins are displayed in Table 4.11. We see again that a pro-social leader has a positive effect on group contribution. Of more interest to us is the optimal game in terms of maximizing group contributions.

Table 4.10: Simulation by leader's type

Leader SVO	Pro-social Followers	Treatments	Nr		Sc		Sf	
			Group	Leader	Group	Leader	Group	Leader
Pro-self	0	Cheaptalk	9.98	2.88	10.56	3.01	10.04	3.13
Pro-self	1	Cheaptalk	9.92	2.90	10.60	2.86	11.06	3.54
Pro-self	2	Cheaptalk	10.97	2.88	11.43	2.89	11.34	3.42
Pro-self	3	Cheaptalk	11.87	2.88	12.18	2.91	11.62	3.35
Pro-self	0	Hide	9.56	2.68	-	-	9.94	3.64
Pro-self	1	Hide	12.16	3.20	-	-	11.26	3.65
Pro-self	2	Hide	12.19	3.06	-	-	12.19	3.65
Pro-self	3	Hide	12.38	2.98	-	-	12.90	3.64
Pro-self	0	Truth	11.63	3.45	-	-	9.28	2.94
Pro-self	1	Truth	12.95	3.45	-	-	10.65	2.93
Pro-self	2	Truth	13.43	3.45	-	-	10.78	2.92
Pro-self	3	Truth	13.94	3.45	-	-	10.92	2.92
Pro-social	0	Cheaptalk	13.47	4.27	13.57	4	13	4
Pro-social	1	Cheaptalk	12.17	4.18	14.09	3.96	12.14	4
Pro-social	2	Cheaptalk	13.25	4.12	14.88	3.94	12.51	4
Pro-social	3	Cheaptalk	14.27	4.08	15.68	3.92	12.84	4
Pro-social	0	Hide	9.53	2.43	13.58	3	-	-
Pro-social	1	Hide	12.78	3.31	14.01	3.50	-	-
Pro-social	2	Hide	12.89	3.30	14.49	3.68	-	-
Pro-social	3	Hide	13.21	3.29	14.98	3.77	-	-
Pro-social	0	Truth	10.5	3	16.79	5	-	-
Pro-social	1	Truth	14.55	4.13	16.45	4.61	-	-
Pro-social	2	Truth	15.13	4.13	16.59	4.44	-	-
Pro-social	3	Truth	15.72	4.13	16.89	4.34	-	-

If the leader is pro-social and sends message ‘Sc’ or ‘Nr’ we see that group contributions are highest in the Truth game compared to the Cheap-talk and Hide games. This reiterates the earlier finding that followers appear more responsive to an ‘Sc’ message in the Truth game than Hide game. Aggregate contributions will depend on q - the probability of message ‘Sc’ in the Truth game. Note, however, that group contributions in the event of message ‘Nr’ in the Truth game are higher than those of ‘Sc’ in the other two games. To illustrate, compare a setting with 2 pro-social followers. In the Cheap-talk and Hide games the leader can send message ‘Sc’ and group contributions are (reading Table 4.10) predicted as 14.88 and 14.49 respectively. In the Truth game predicted contributions are $16.59(1-q)+15.13q$. In our experiment, where $q=0.7$ this gives 16.15. The Truth game, thus, has highest contributions even though a pro-social leader cannot send message ‘Sc’ for sure. This goes against Hypothesis 4.6 with contributions relatively lower in the Hide game and higher in the Truth game than predicted.

Result 4.11. *If the leader is pro-social, the aggregate group contributions are higher in the Truth game than Hide and Cheap-talk games.*

If the leader is pro-self and reveals type (i.e. sends message ‘Sf’) then there is little discernible difference across games although contributions are lowest in the Truth game. If the leader sends message ‘Nr’ then group contributions are highest in the Truth game. Overall contributions will, thus, critically depend on the message sent. Again, consider a case where 2 followers are pro-social. In the Cheap-talk game, if the leader sends message ‘Sc’ we have predicted group contributions of 11.43. In the Hide game, if the leader sends message ‘Nr’ we have predicted group contributions of 12.19. In the Truth game predicted contributions are $10.78(1-q)+13.43q$. If $q=0.7$ this gives 11.58. Group contributions are, therefore, marginally higher in the Hide game and lowest in the Cheap-talk game. This goes against Hypothesis 4.7. Overall, however, we see relatively little difference across treatments.

Table 4.11: Regression on group contribution simulation by leader types

	(1)		(2)	
	Total group contribution		Margins	
Leader's Contribution	2.688***	(2962.74)		
Pro-social Followers=1	1.282***	(45.34)		
Pro-social Followers=2	1.761***	(145.88)		
Pro-social Followers=3	2.233***	(209.55)		
Cheaptalk	0.287***	(45.40)	12.87***	(5609.72)
Truth	-0.258***	(-73.62)	12.87***	(8161.92)
Pro-social Followers=1 × Cheaptalk	-0.914***	(-28.42)		
Pro-social Followers=1 × Truth	0.0794***	(2.58)		
Pro-social Followers=2 × Cheaptalk	-0.578***	(-39.62)		
Pro-social Followers=2 × Truth	-0.177***	(-12.93)		
Pro-social Followers=3 × Cheaptalk	-0.317***	(-24.19)		
Pro-social Followers=3 × Truth	-0.404***	(-32.28)		
Pro-social Leader	2.295***	(69.77)	13.62***	(2280.11)
Pro-social Followers=1 × Pro-social Leader	-0.874***	(-19.48)		
Pro-social Followers=2 × Pro-social Leader	-1.355***	(-38.23)		
Pro-social Followers=3 × Pro-social Leader	-1.574***	(-45.04)		
Cheaptalk × Pro-social Leader	-1.981***	(-52.03)	13.09***	(1964.51)
Truth × Pro-social Leader	-0.572***	(-14.81)	13.69***	(1848.04)
Pro-social Followers=1 × Cheaptalk × Pro-social Leader	1.194***	(23.35)		
Pro-social Followers=1 × Truth × Pro-social Leader	0.153***	(3.01)		
Pro-social Followers=2 × Cheaptalk × Pro-social Leader	1.691***	(41.43)		
Pro-social Followers=2 × Truth × Pro-social Leader	0.983***	(23.74)		
Pro-social Followers=3 × Cheaptalk × Pro-social Leader	2.033***	(50.50)		
Pro-social Followers=3 × Truth × Pro-social	1.509***	(36.78)		
Hide			13.02***	(5357.15)
Pro-self Leader			12.34***	(6742.38)
Cheaptalk × Pro-self Leader			12.52***	(3891.39)
Hide × Pro-self Leader			12.56***	(2506.06)
Hide × Pro-social Leader			13.96***	(1163.34)
Truth × Pro-self Leader			12.14***	(4404.26)
Constant	1.843***	(491.53)		
Observations	5311877		5311877	

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 4.12. *If the leader is pro-self, aggregate contributions are marginally higher in the Hide than Cheap-talk and Truth games.*

Recall that, aggregating across types, total contributions were highest in the Cheap-talk game (13.89), next highest in the Truth game (12.68) and lowest in the Hide game (11.28). Given Results 4.11 and 4.12, this aggregate finding is driven by a range of competing forces. The main reason why Cheap-talk comes out top is that 71% of leaders send message 'Sc' compared to only 43% in the Truth game and 50% in the Hide game. While an 'Sc' message receives less reciprocation in the Cheap-talk game (see Result 9), the effect is not enough to offset the higher proportion of 'Sc' messages. Thus, Cheap-talk results in the highest contributions. In the Truth game we set $q = 0.7$ meaning that 30% of pro-social leaders did not send message 'Sc'. If we had set $q = 1$ then, ceteris paribus, we would have seen around 60% of leaders send message 'Sc' (see Table 4.2). This is still less than the 71% observed in the Cheap-talk game. Moreover we would have seen a counter-balancing increase in message 'Sf'. The finding that contributions are highest in the Cheap-talk game would seem, therefore, robust to changes in q.

4.6 Conclusion

In the present chapter, we investigated the effect of leaders' SVO types on followers' cooperative behaviour and group total contribution in a public good game. Moreover, we investigated whether leaders choose to reveal or 'hide' their SVO type from followers. There were three treatments in the experiment – Cheap-talk, Hide and Truth. In the Cheap-talk game leaders had the freedom to choose tell followers that the leader in their group is a pro-social/pro-self or that type is not revealed. This means that a pro-self leader has the opportunity to say they are pro-social. In the Hide game, leaders could choose to send a message about their true SVO type or hide their type to the followers in the group. This means a pro-self leader could 'hide' but not 'lie' about type. In the Truth game the computer randomly decided whether to reveal or not reveal leader type.

We found that, when it comes to the messages sent by leaders, the majority of pro-social leaders choose to send a message revealing their true type. However, the majority of pro-self leaders prefer not to disclose their true type. Instead, they will reveal themselves as pro-social or just hide their SVO type. We also found that pro-social leaders and followers contribute significantly more than pro-self leaders and followers. Furthermore, we found high levels of conditional cooperation with the contribution of followers rising in tandem with the contribution of the leader.

In terms of the effect of messages, we find that contributions are higher if the leader sends a pro-social message, for a fixed leader contribution, than if the message does not reveal type or is pro-self. Furthermore, we observe that pro-selfs appear to regard leaders of unknown type as if they were pro-self, while pro-socials appear to regard them as pro-social. We find that contributions are higher, ceteris paribus, in the Truth game than the other two games if the leader sends a pro-social message, yet the treatment effects are insignificant. It appears, therefore, that followers react to the message rather than the credibility of the message. This is consistent with the notion that the message changes beliefs which then changes contributions.

Finally we can turn to the question of which game is best for overall efficiency. We find that the Cheap-talk game comes out top in terms of average contributions to the public good. The basic logic behind this result is that in the Cheap-talk game a large proportion of leaders send the message that they are pro-social and this then increases contributions. It seems, therefore, that it is enough for the leader to be able to claim they are pro-social to increase efficiency. This finding needs, however, to be qualified relative to the type of the leader. If the leader is pro-social then contributions are highest in the Truth game. This is because the pro-social leader can credibly signal their type. The optimal game or mechanism to use, thus, depends on the likely type of the leader. If the leader is randomly drawn from the population (and, therefore, has a good chance to be pro-self) the Cheap-talk game appears optimal. If, however, the leader is selected and likely to be pro-social the Truth game appears optimal.

Appendix A

A.1 Chapter Two

A.1.1 Experiment Instructions

[Instructions in English (translated from Chinese)]

Welcome!

This is a decision-making experiment with several rounds. You are randomly assigned into a group of 4 members, and will remain in the same group during the experiment. Every member of the group is given an endowment of 20 tokens per round. In every round, all members of the group must simultaneously and anonymously decide how many points to contribute to a Group Account. You can contribute any (whole number) amount between 0 and 20.

Once everyone has made their choice we will calculate the total amount of points allocated to the Group Account in your group. Total points in the Group Account will be multiplied by 1.6 and then divided equally between all 4 members in your group. Your earnings in a round will be equal to your return from the Group Account plus the remaining tokens from your endowment. Earnings = 20 - contribution to Group Account + 1.6 x Total contributions to Group Account / 4.

For example: If you contribute 8 points and the total points in the public account is 30 then you earn (20-8) + 30*1.6/4 = 24 points in this round.

Payment

After all the rounds, the value of the accumulated points you eventually earned from all rounds will be converted into real world currency (£) to be paid to you, in addition to a £5 show-up fee.

For example: If the sum of points you get from all decision-making rounds is 260 points, you will be paid £2.60 plus £5.00 in the end of the experiment.

A preliminary task

Before we begin the group project and the main part of the experiment we will do a preliminary task.

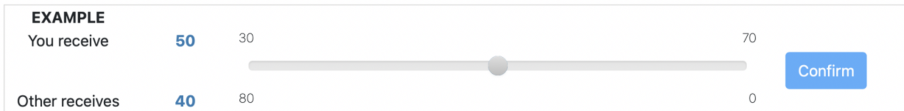
Now, imagine that you are randomly matched with another person in the room. And you have to decide how to allocate tokens between you and that person. We will give you six scenarios to consider (on the next page). The example below shows how to make your choice.

Try to move the little slider along the scale to make your allocation depending on your choice. Confirm it when you are happy about the decision. (Note that once you have confirmed it you will be unable to change it.)

Note: In each scenario, the blue numbers indicate the final allocation according to your decision, while the grey numbers indicate the allocation at two endpoints.

The blue numbers will change as you move the slider but the grey ones are fixed.

For example: in the allocation below, you will receive 50 tokens and leave 40 tokens to the other person.



Preliminary task: Results

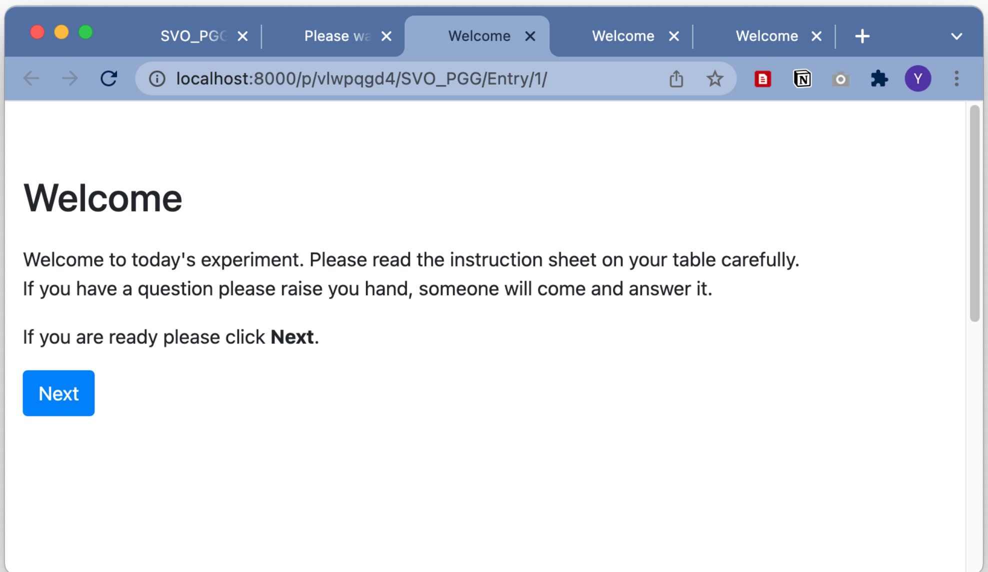
The task you have just performed is a standard psychological test used to classify people as either pro-self or pro-social.

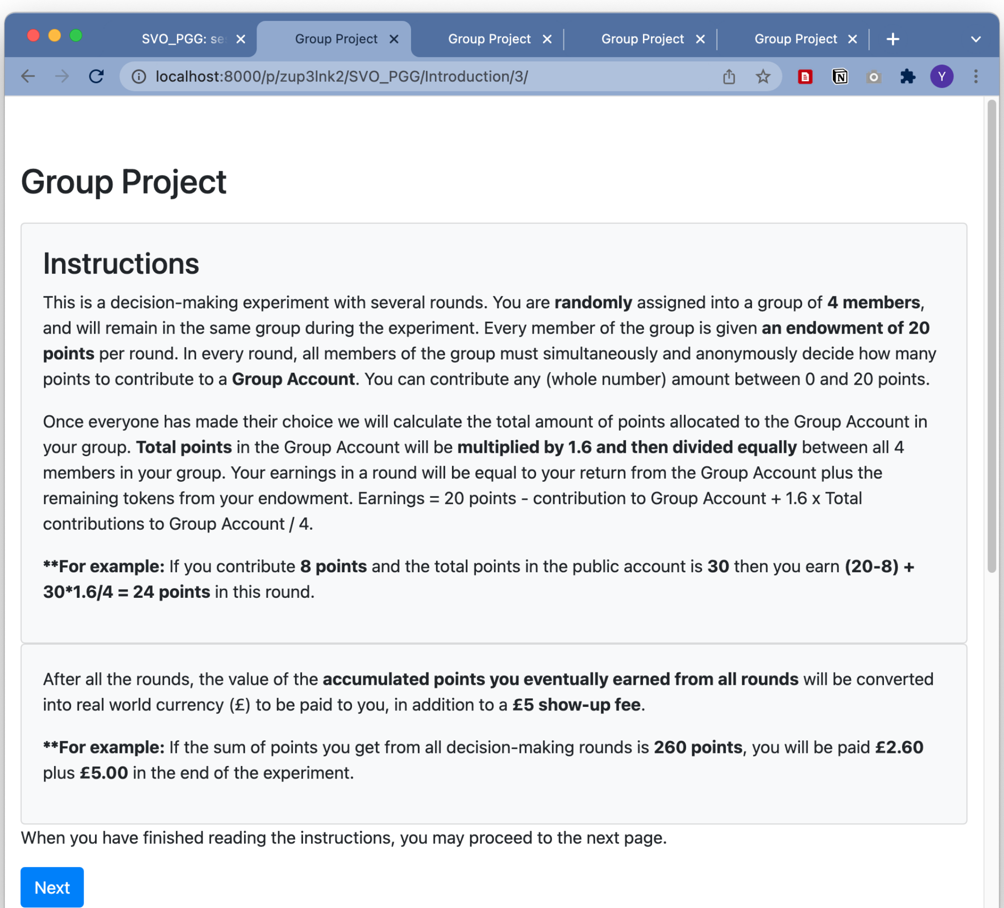
A person's behaviour is pro-social if it benefits others while it is pro-self if it benefits the person themselves.

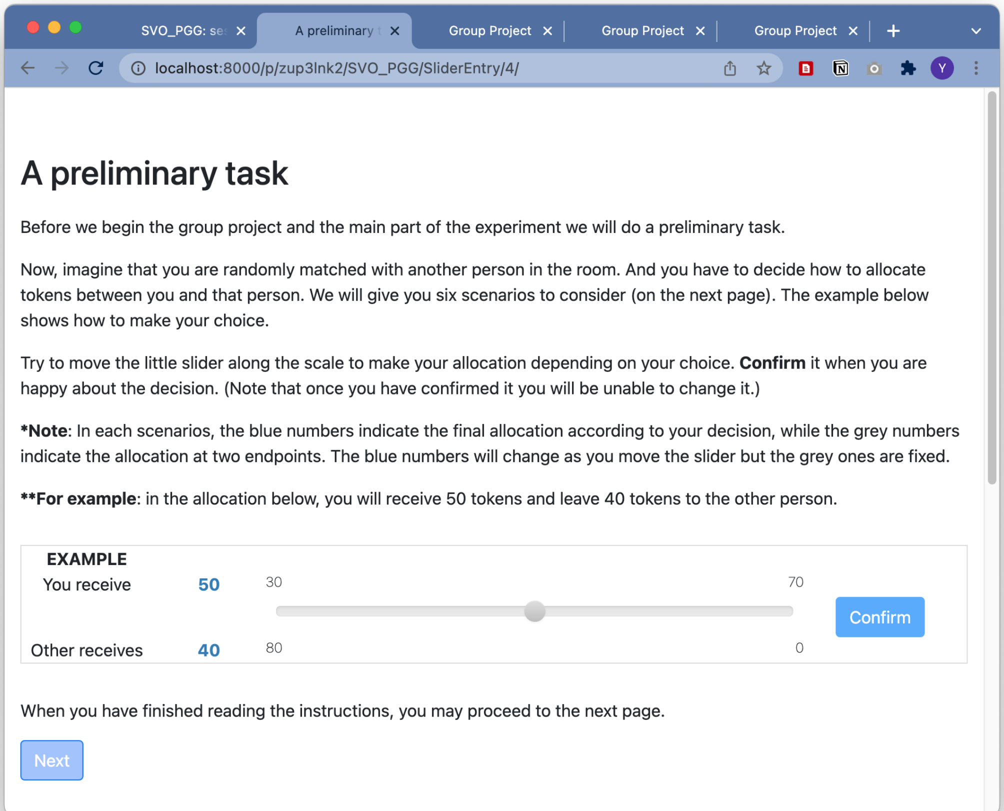
Based on the choices made, you and everyone else in your group has been classified as either pro-social or pro-self.

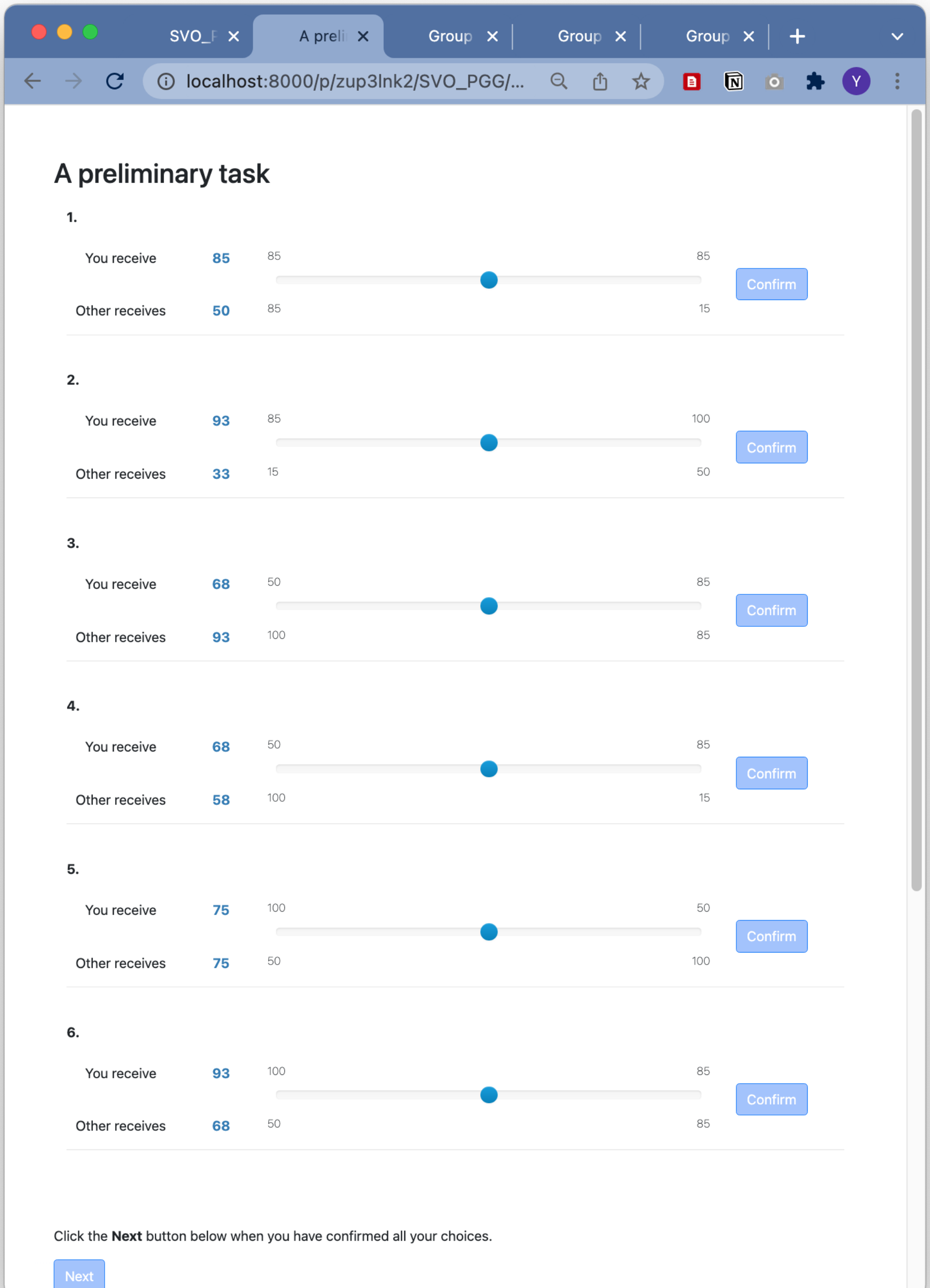
Note: The classification of you and others in your group may be shared anonymously amongst the group members in the remainder of the experiment. You will be told if and when this happens.

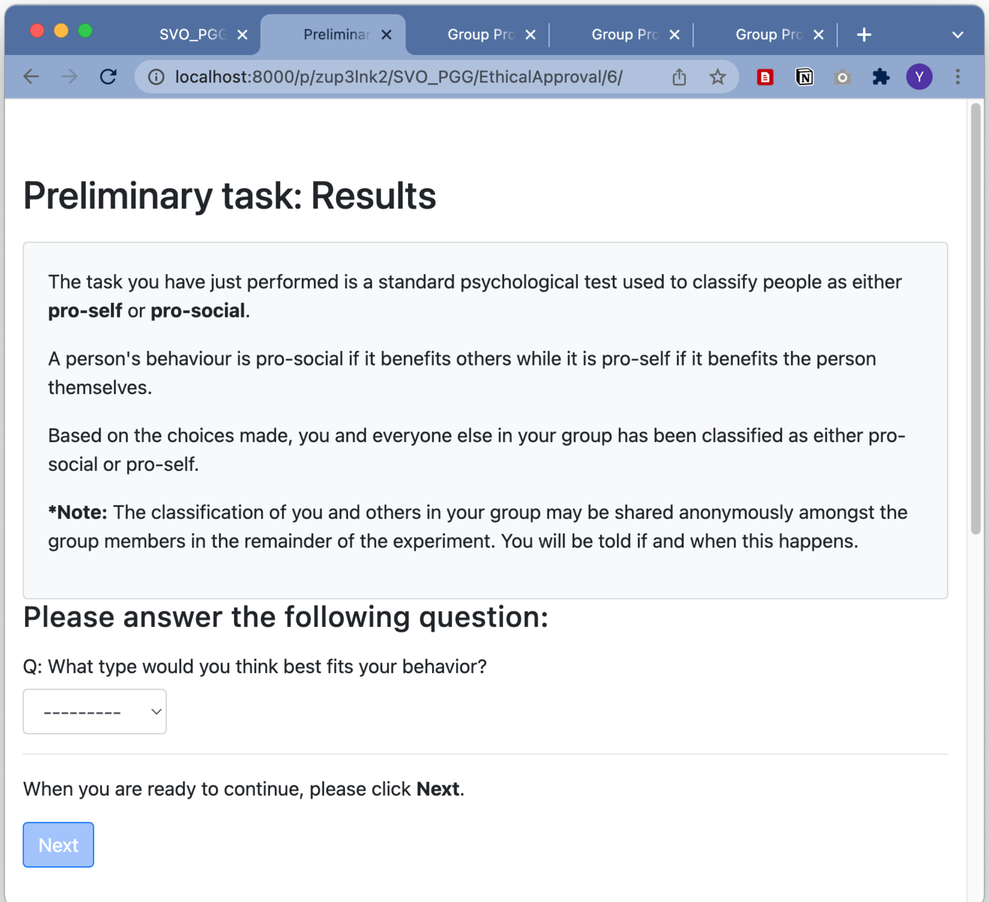
A.1.2 Experiment Main Pages Screenshots



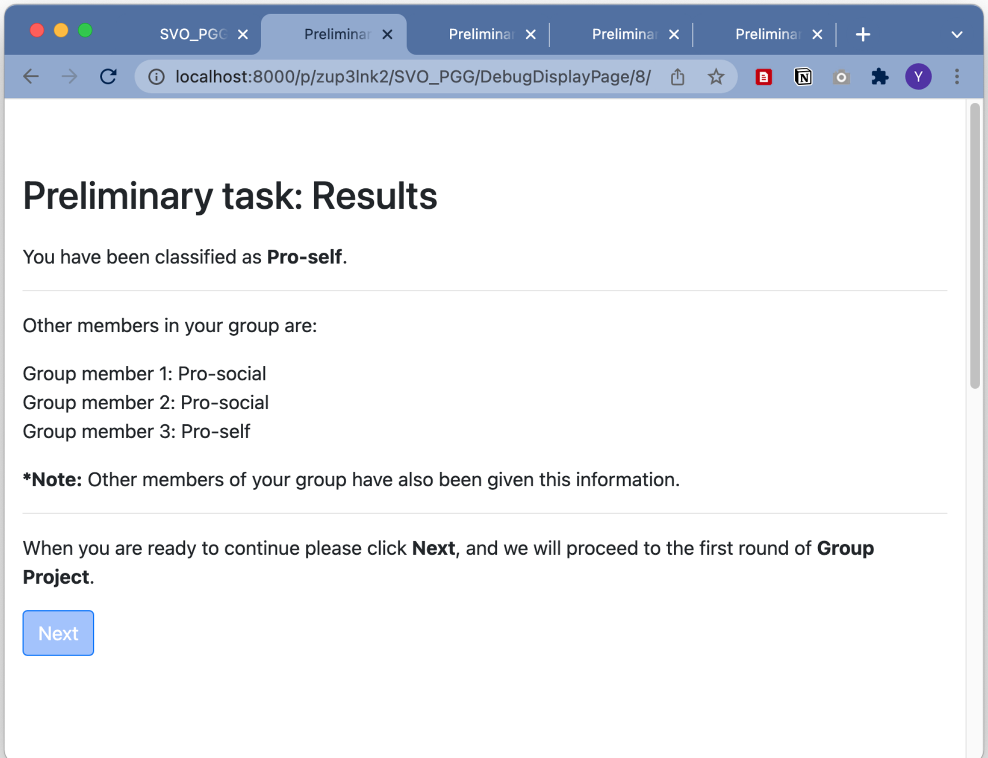


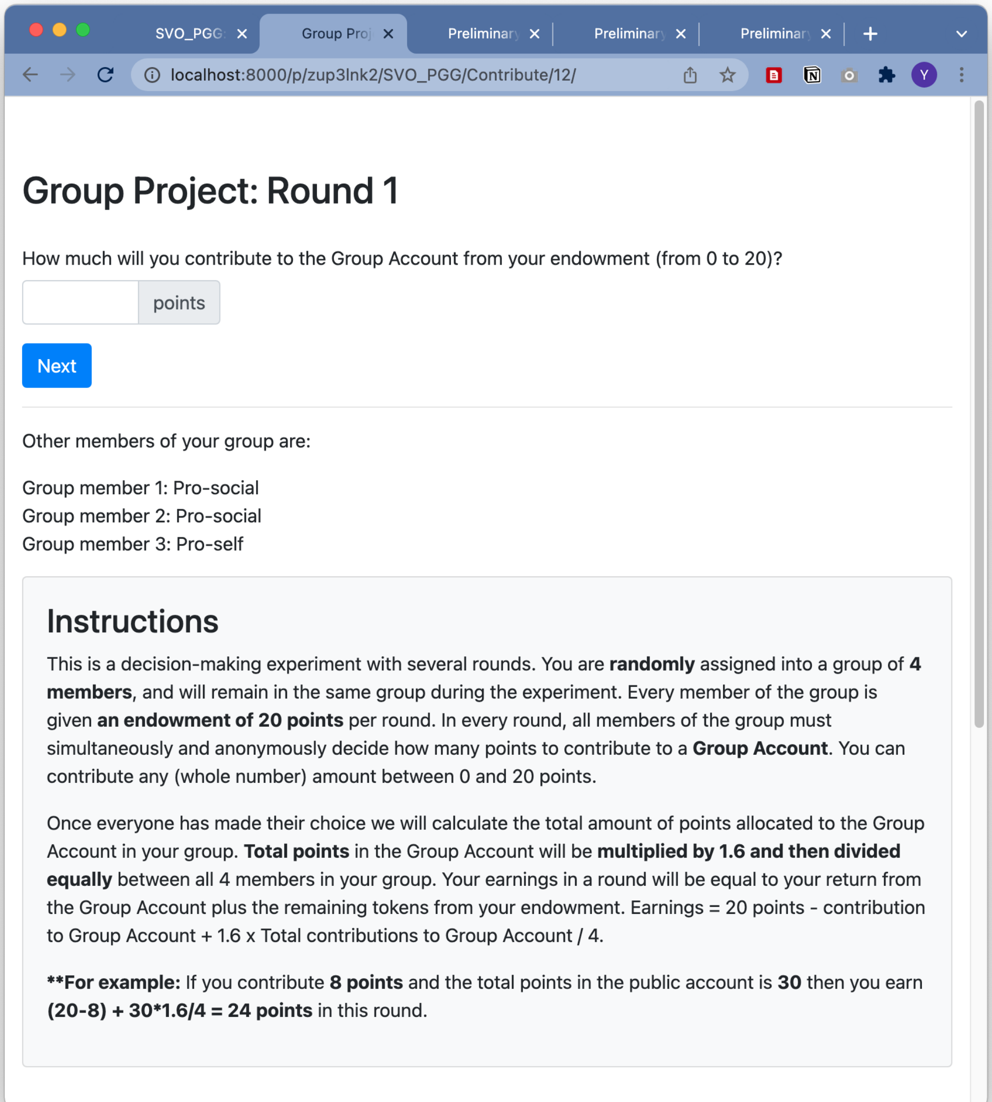

SVO_PGG: se
A preliminary t
Group Project
Group Project
Group Project
localhost:8000/p/zup3lnk2/SVO_PGG/SliderEntry/4/
A preliminary task
Before we begin the group project and the main part of the experiment we will do a preliminary task.
Now, imagine that you are randomly matched with another person in the room. And you have to decide how to allocate tokens between you and that person. We will give you six scenarios to consider (on the next page). The example below shows how to make your choice.
Try to move the little slider along the scale to make your allocation depending on your choice. Confirm it when you are happy about the decision. (Note that once you have confirmed it you will be unable to change it.)
*Note: In each scenarios, the blue numbers indicate the final allocation according to your decision, while the grey numbers indicate the allocation at two endpoints. The blue numbers will change as you move the slider but the grey ones are fixed.
**For example: in the allocation below, you will receive 50 tokens and leave 40 tokens to the other person.
EXAMPLE
You receive
50
30
70
Confirm
Other receives
40
80
0
When you have finished reading the instructions, you may proceed to the next page.
Next

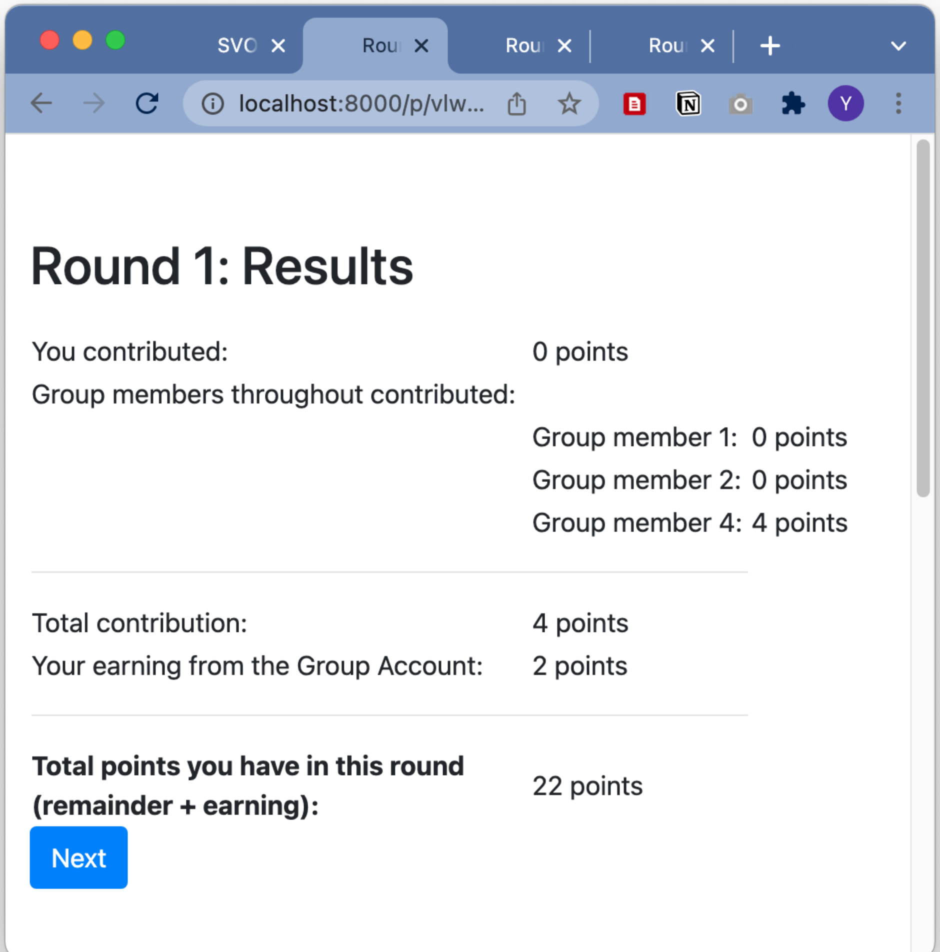

SVO_P
A preli
Group
Group
Group
localhost:8000/p/zup3lnk2/SVO_PGG/...
A preliminary task
1.
You receive 85
85 85
Other receives 50
85 15
Confirm
2.
You receive 93
85 100
Other receives 33
15 50
Confirm
3.
You receive 68
50 85
Other receives 93
100 85
Confirm
4.
You receive 68
50 85
Other receives 58
100 15
Confirm
5.
You receive 75
100 50
Other receives 75
50 100
Confirm
6.
You receive 93
100 85
Other receives 68
50 85
Confirm
Click the Next button below when you have confirmed all your choices.
Next

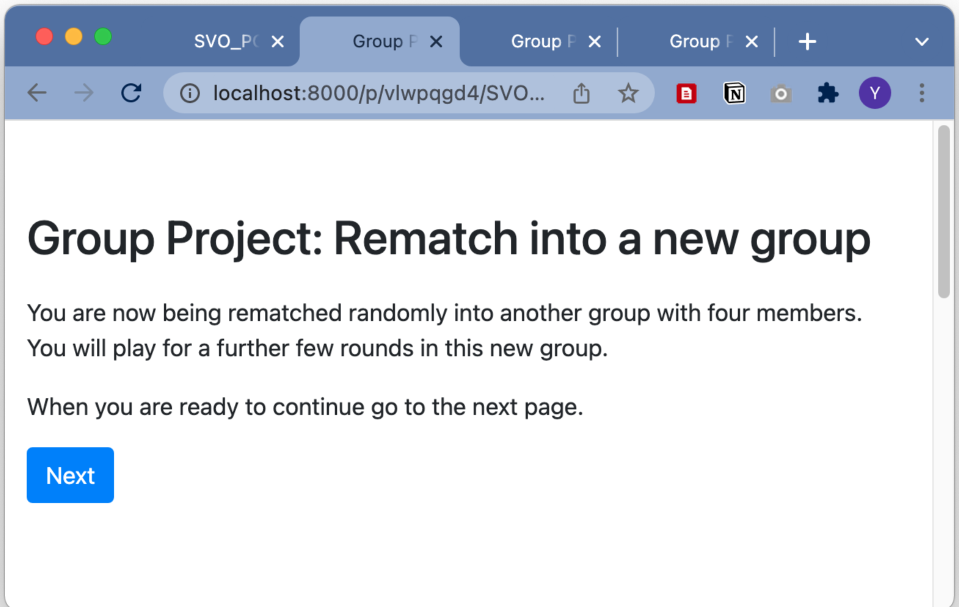

SVO_PGG
Preliminar
Group Pro
Group Pro
Group Pro
localhost:8000/p/zup3lnk2/SVO_PGG/EthicalApproval/6/
Preliminary task: Results
The task you have just performed is a standard psychological test used to classify people as either pro-self or pro-social.
A person's behaviour is pro-social if it benefits others while it is pro-self if it benefits the person themselves.
Based on the choices made, you and everyone else in your group has been classified as either pro-social or pro-self.
*Note: The classification of you and others in your group may be shared anonymously amongst the group members in the remainder of the experiment. You will be told if and when this happens.
Please answer the following question:
Q: What type would you think best fits your behavior?

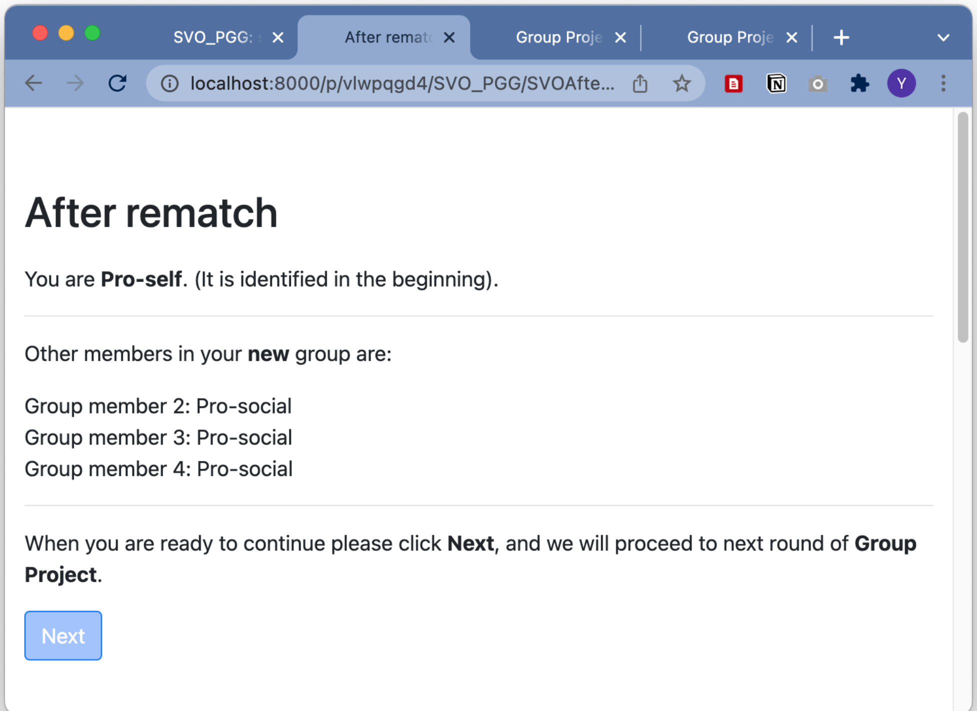
When you are ready to continue, please click Next.
Next

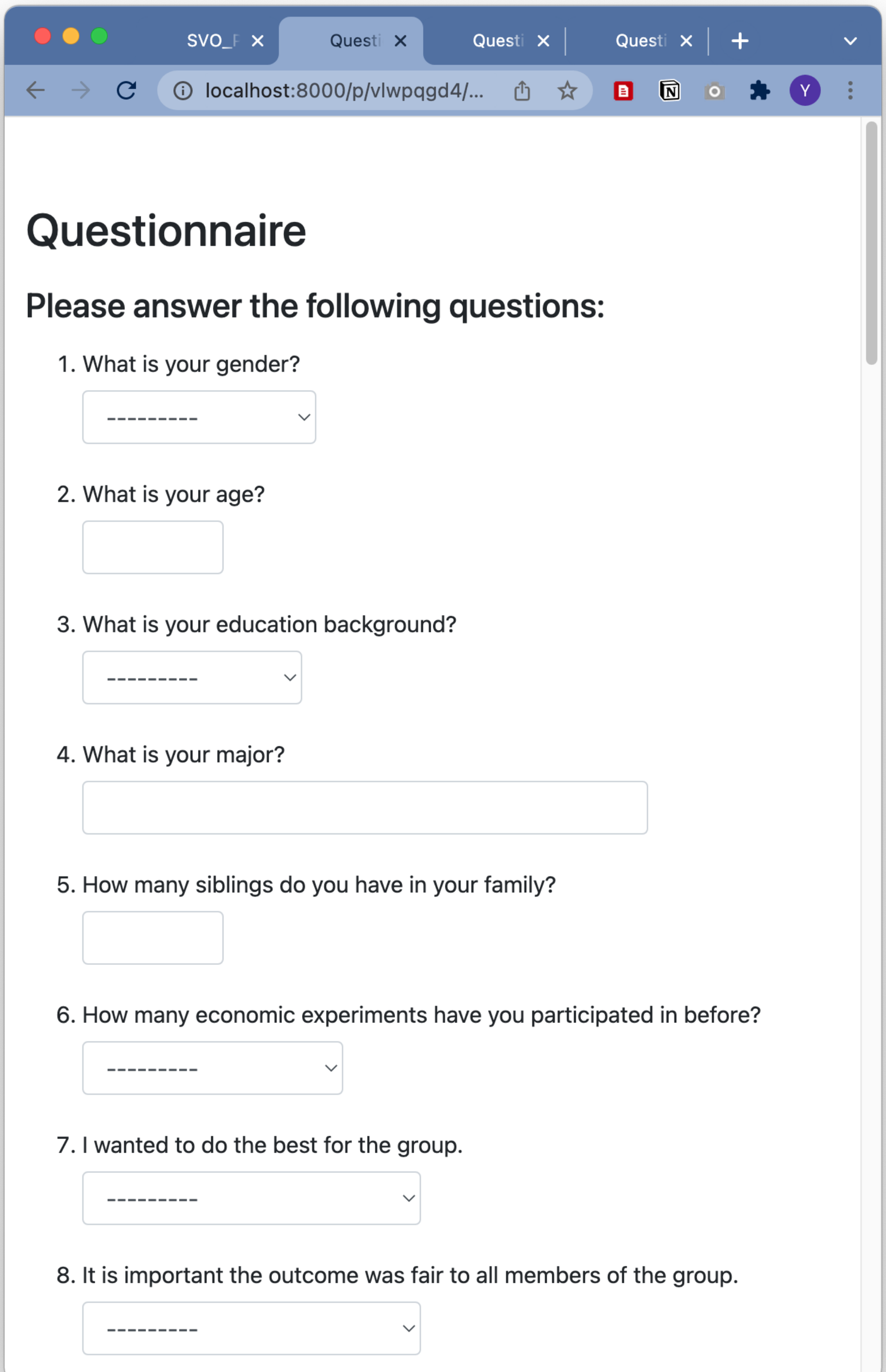

SVO_PGG
Preliminar
Preliminar
Preliminar
Preliminar
localhost:8000/p/zup3lnk2/SVO_PGG/DebugDisplayPage/8/
Preliminary task: Results
You have been classified as Pro-self.
Other members in your group are:
Group member 1: Pro-social
Group member 2: Pro-social
Group member 3: Pro-self
*Note: Other members of your group have also been given this information.
When you are ready to continue please click Next, and we will proceed to the first round of Group Project.
Next


SVO_PGG
Group Proj
Preliminary
Preliminary
Preliminary
localhost:8000/p/zup3lnk2/SVO_PGG/Contribute/12/
Group Project: Round 1
How much will you contribute to the Group Account from your endowment (from 0 to 20)?
points
Next
Other members of your group are:
Group member 1: Pro-social
Group member 2: Pro-social
Group member 3: Pro-self
Instructions
This is a decision-making experiment with several rounds. You are randomly assigned into a group of 4 members, and will remain in the same group during the experiment. Every member of the group is given an endowment of 20 points per round. In every round, all members of the group must simultaneously and anonymously decide how many points to contribute to a Group Account. You can contribute any (whole number) amount between 0 and 20 points.
Once everyone has made their choice we will calculate the total amount of points allocated to the Group Account in your group. Total points in the Group Account will be multiplied by 1.6 and then divided equally between all 4 members in your group. Your earnings in a round will be equal to your return from the Group Account plus the remaining tokens from your endowment. Earnings = 20 points - contribution to Group Account + 1.6 x Total contributions to Group Account / 4.
**For example: If you contribute 8 points and the total points in the public account is 30 then you earn (20-8) + 30*1.6/4 = 24 points in this round.


SVO
Rou
Rou
Rou
localhost:8000/p/vlw...
Round 1: Results
You contributed: 0 points
Group members throughout contributed:
Group member 1: 0 points
Group member 2: 0 points
Group member 4: 4 points
Total contribution: 4 points
Your earning from the Group Account: 2 points
Total points you have in this round (remainder + earning): 22 points
Next


SVO_P
Group P
Group P
Group P
localhost:8000/p/vlwpqgd4/SVO...
Group Project: Rematch into a new group
You are now being rematched randomly into another group with four members.
You will play for a further few rounds in this new group.
When you are ready to continue go to the next page.
Next


SVO_PGG:
After remat
Group Proje
Group Proje
localhost:8000/p/vlwpqgd4/SVO_PGG/SVOAfte...
After rematch
You are Pro-self. (It is identified in the beginning).
Other members in your new group are:
Group member 2: Pro-social
Group member 3: Pro-social
Group member 4: Pro-social
When you are ready to continue please click Next, and we will proceed to next round of Group Project.
Next



Questionnaire

Please answer the following questions:

1. What is your gender?

   ----

2. What is your age?

3. What is your education background?

   ----

4. What is your major?

5. How many siblings do you have in your family?

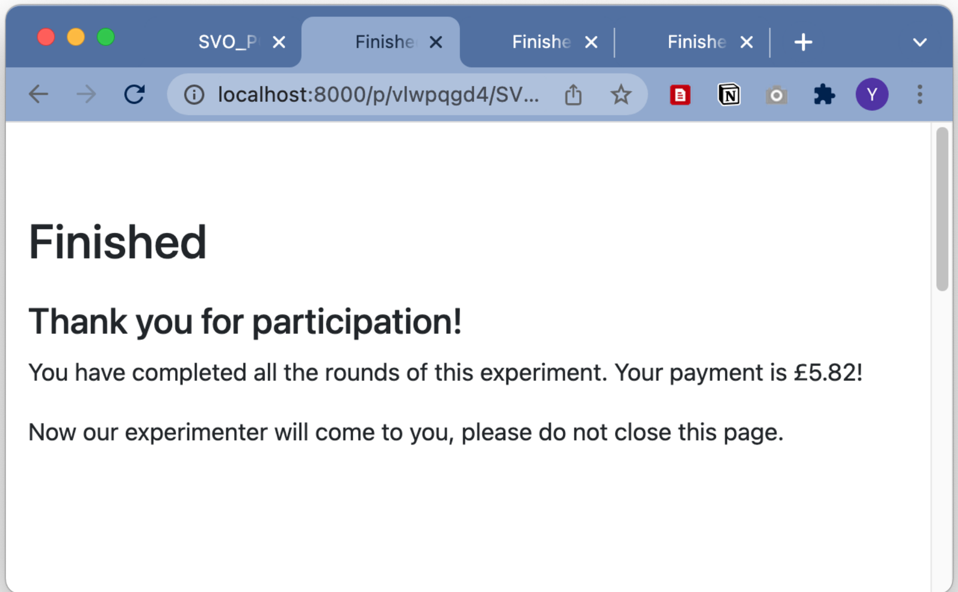
6. How many economic experiments have you participated in before?

   ----

7. I wanted to do the best for the group.

   ----

8. It is important the outcome was fair to all members of the group.



A.2 Chapter Three

A.2.1 Experiment Instructions

Welcome!

Thank you for taking an interest in our research study. Before you continue it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information before clicking your consent to take part. Also please note that at the end of the survey there is a link to follow to process payment.

What is the purpose of the study?

The primary aim of this study is to better understand cooperation in groups. We will also ask you some questions about your use of cyber security measures.[1]

What does the study involve?

The survey should take approximately 15 minutes to complete. The survey will consist of tick box questions that elicit your views and preferences. Your answers will be treated confidentially and the information you provide will be kept fully anonymous in any research outputs/publications.

Points?

In the first part of the survey you will be involved in a choice decision involving points. At the end of the experiment you will be given a bonus payment based on the points you earn where 1 point = 5 pence.

Who is doing the research?

The study is being conducted by Yidan Chai from the University of Kent and Professor Edward Cartwright from De Montfort University.

Taking part?

Your participation in the survey is entirely voluntary, and you can opt-out at any stage by closing and exiting the browser. If you are happy to take part then please follow the survey until the end and then click on the link to collect payment.

What happens to the information I provide?

[1] The study is combined with another study of cyber-security which is not described in this project.

Data from the survey will be held securely on a password-protected file on a password-protected account. Before 01/11/2020 the data will be fully anonymised and any non-anonymous data will be deleted. The anonymous version of the data will be put on a research data repository and used in the production of formal research outputs (e.g. journal articles, conference papers, theses, and reports).

Making a complaint

For further information, or if you have any queries, please contact the lead researcher Yidan Chai (yc290@kent.ac.uk). For further information please contact Yidan Chai (yc290@kent.ac.uk)

[Consent form]

I have read and understood the above information. Yes

I understand that, because my answers will be fully anonymised, it will not be possible to withdraw them from the study once I have completed the survey. Yes

I agree to take part in this survey. Yes

I confirm that I am aged 18 or over. Yes

Please enter your Prolific ID in the box below. This will allow us to arrange any additional payments to you:

[Participants enter their Prolific ID in here]

If you are ready please click Next.

Group Project

We will begin with a decision making experiment in which you can earn points which will be converted into a bonus payment.

You are **randomly** assigned into a group of **4 members**. Everyone in the group is given **an endowment of 20 points.**

You will need to make two decisions sequentially.

1. In the first decision stage, you decide how many tokens to invest into a **Group Account** from your endowment. You can contribute any (whole number) amount between 0 and 20 points.

2. In the second decision stage, you fill out an **"investment table"**. In the investment table you have to indicate for each possible investment of one other group member how many tokens you want to invest in the Group Account. You can condition your investment on the investment of one other group member.

Payment

After all participants have made the two decisions:

One group member will be randomly selected by the system, his/her investment in the first stage will be allocated into the Group Account.

For **other group members**, the decisions in the investment table will be allocated into the Group Account. Therefore, either of the two decisions you will make could be the one that determines your payoff.

Earnings from Group Account:

Once the investments are determined, total amount of tokens allocated to the Group Account in your group will be **multiplied by 1.6, and then divided equally** between all 4 members in your group.

Your earnings will be: Earnings = 20 points - your contribution to Group

Account + 1.6 x Total amount in Group Account / 4.

***For example: If you contributed 8 and the total points in the Group Account is 30, then you will earn (20-8) + 30*1.6/4 = 24 points.*

Final Payoff:

At the end of the experiment, the value of the **points you earned** will be converted into real world currency (£) to be paid to you, in addition to a **participation fee**.

***For example: If the earnings you get from the Group Account is 30 points, you will be paid £1.5 plus participation fee in the end of the experiment.*

A preliminary task

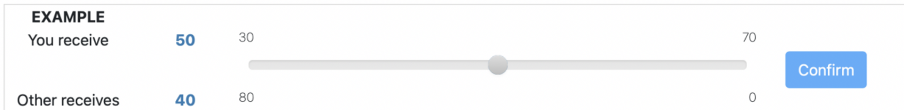
Before begin the group project, let's do a preliminary task first.

Now, imagine that you are randomly matched with another person in the study. And you have to decide how to allocate points between you and that person. We will give you six scenarios to consider (on the next page).

When making the decisions, move the little slider along the scale to make your allocation. **Confirm** it when you are happy about the decision. (Note that once you have confirmed it you will be unable to change it.)

In each scenario, the **blue** numbers indicate the final allocation according to your decision, while the **grey** numbers indicate the allocation at two endpoints. The blue numbers will change as you moving the slider but the grey ones are fixed.

*The **example graph** below shows one decision-making situation: you will receive 50 points and leave 40 points to the other person. (Keep in mind this is just an example.)*



When you have finished reading the instructions, you may proceed to the next page.

A preliminary task

[Participants complete the SVO task by moving the sliders]

Preliminary task: Results

The task you have just performed is a standard psychological test used to classify people as either **pro-social** or **pro-self**:

A person's behaviour is pro-social if it benefits others while it is pro-self if it benefits the person themselves.

Based on the choices made, you and everyone else in your group has been classified as either pro-social or pro-self.

**Note: Your classification will not be shared with others, and you will not be told the classification of other group members.*

Please answer the following question:

Q: What type would you think best fits your behavior? Pro-self/Pro-social

Preliminary task: Results

[Treatment Binary] You have been classified as Pro-social/Pro-self.

[Treatment Scale] On the scale of the measure, your results are indicated as the slider block locating: Very pro-self/ Very pro-social

***Note: the more closer to the right-hand endpoint, the more pro-social your results are.

[Treatment No info] [No information of SVO results provided]

When you are ready to continue please click **Next**, and we will proceed to the first decision stage of **Group Project**.

Group Project: first decision stage

We will now return to the decision making experiment. If you need reminding of the instructions then they are repeated below.

How much will you contribute to the Group Account from your endowment (from 0 to 20)?

[Participants enter a number between 0 to 20]

Group Project: second decision stage

Now, imagine that there is a member in your group will invest first. Then you and other two members will invest after him/her.

***Recall that you will get earnings from the Group Account depending on the total amount tokens in the account. The tokens you did not invest will be another part of your final payoff.*

The decisions made in the table below would affect your final payoff.

Please indicate how much you would invest in different scenarios:

[Participants complete the table by entering 21 numbers indicating conditional contributions given other member contributes 0 to 20.]

Questionnaire

Please answer the following questions:

What is your gender? Male/Female/Other/Prefer not to say

What is your age? 18-25/ 26-30/ 31-40/ 41-50/ 51-60/ 61-70/ Over 70

What is your current employment status? Employed (full-time)/ Employed (part-time)/ Self-employed/ Student/ Retired/ Unemployed (seeking employment)/ Unemployed (not currently seeking employment)

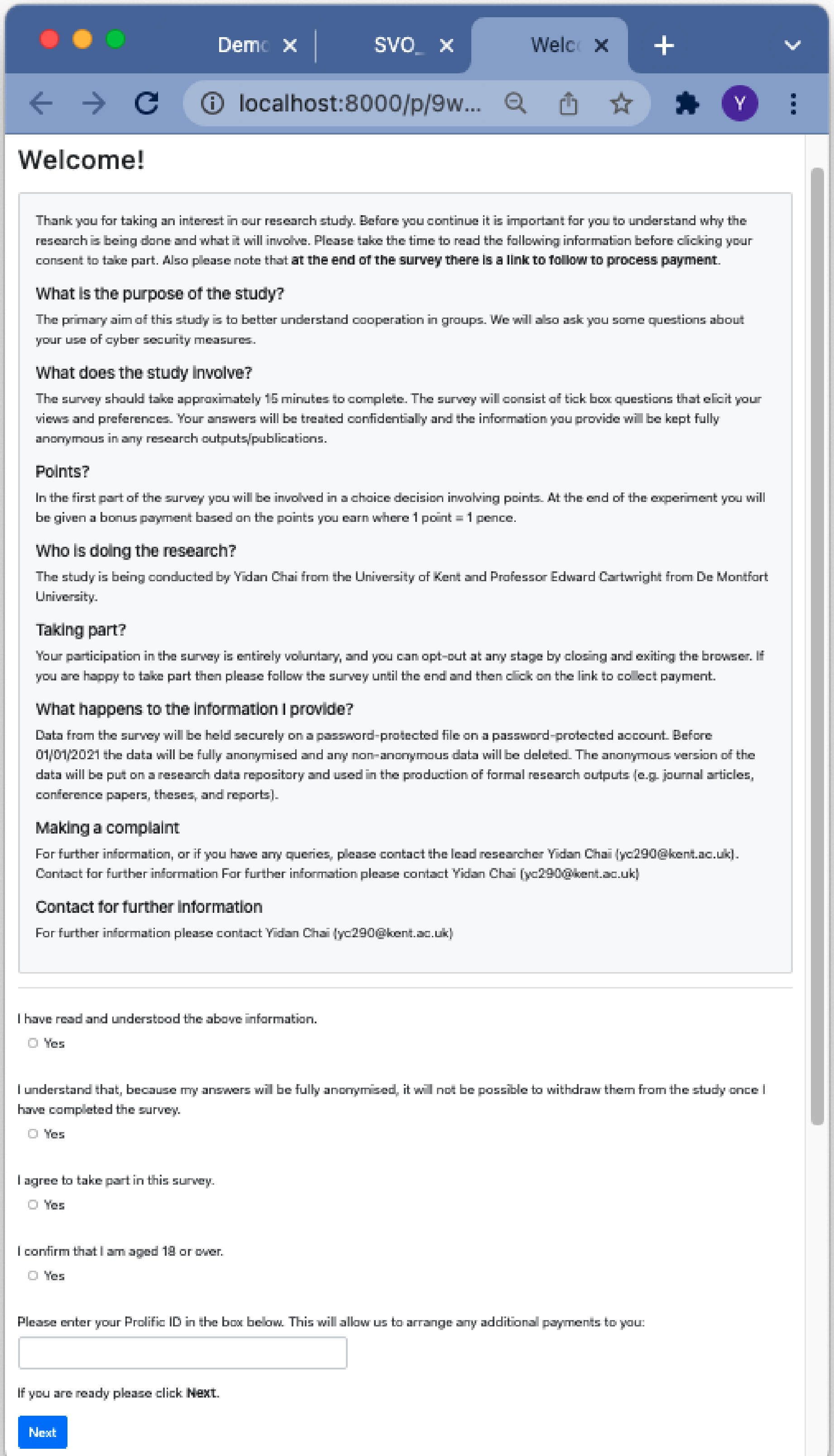
What is your annual household income? Below £10,000/ £10,000 to £30,000/ £30,000 to £60,000/ Above £60,000

What is your highest level of education? GCSE/O-levels / A-Levels/ College Degree/ University Undergraduate Degree/ Postgraduate Degree/ No formal qualifications

[Then a series of unrelated questions about cyber-security behavior.]

A.2.2 Experiment Main Pages Screenshots

Here we only presented main pages which are different from experiment 1. Therefore, pages, such as SVO instructions, test and beliefs are not included here. Those pages are the same - in terms of structures and order - as what we did in experiment 1.



Demo SVO_ Welc

localhost:8000/p/9w...

Welcome!

Thank you for taking an interest in our research study. Before you continue it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information before clicking your consent to take part. Also please note that **at the end of the survey there is a link to follow to process payment**.

What is the purpose of the study?

The primary aim of this study is to better understand cooperation in groups. We will also ask you some questions about your use of cyber security measures.

What does the study involve?

The survey should take approximately 15 minutes to complete. The survey will consist of tick box questions that elicit your views and preferences. Your answers will be treated confidentially and the information you provide will be kept fully anonymous in any research outputs/publications.

Points?

In the first part of the survey you will be involved in a choice decision involving points. At the end of the experiment you will be given a bonus payment based on the points you earn where 1 point = 1 pence.

Who is doing the research?

The study is being conducted by Yidan Chai from the University of Kent and Professor Edward Cartwright from De Montfort University.

Taking part?

Your participation in the survey is entirely voluntary, and you can opt-out at any stage by closing and exiting the browser. If you are happy to take part then please follow the survey until the end and then click on the link to collect payment.

What happens to the information I provide?

Data from the survey will be held securely on a password-protected file on a password-protected account. Before 01/01/2021 the data will be fully anonymised and any non-anonymous data will be deleted. The anonymous version of the data will be put on a research data repository and used in the production of formal research outputs (e.g. journal articles, conference papers, theses, and reports).

Making a complaint

For further information, or if you have any queries, please contact the lead researcher Yidan Chai (yc290@kent.ac.uk). Contact for further information For further information please contact Yidan Chai (yc290@kent.ac.uk)

Contact for further information

For further information please contact Yidan Chai (yc290@kent.ac.uk)

I have read and understood the above information.

○ Yes

I understand that, because my answers will be fully anonymised, it will not be possible to withdraw them from the study once I have completed the survey.

○ Yes

I agree to take part in this survey.

○ Yes

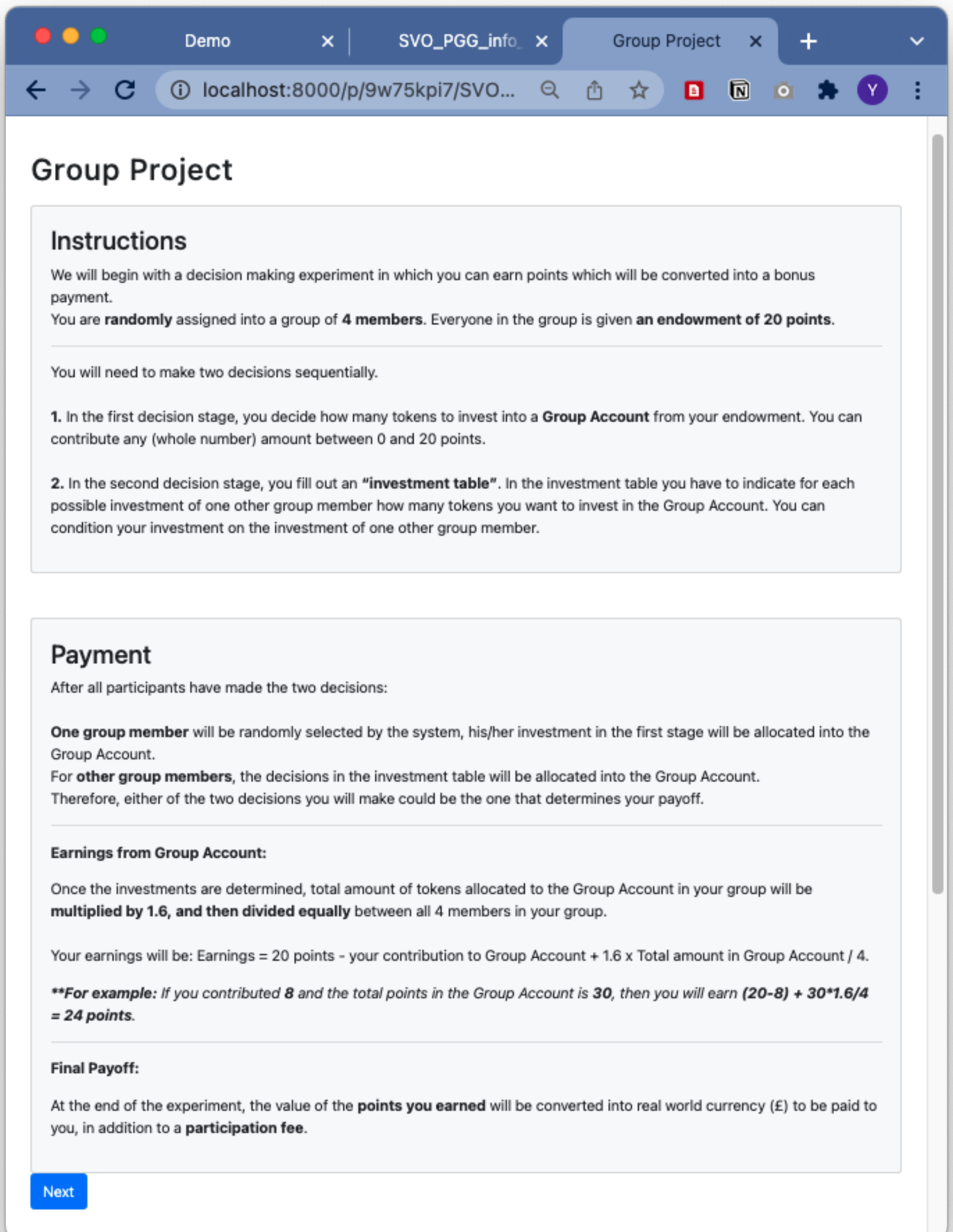
I confirm that I am aged 18 or over.

○ Yes

Please enter your Prolific ID in the box below. This will allow us to arrange any additional payments to you:

If you are ready please click **Next**.

Next



Group Project

Instructions

We will begin with a decision making experiment in which you can earn points which will be converted into a bonus payment.
You are **randomly** assigned into a group of **4 members**. Everyone in the group is given **an endowment of 20 points**.

You will need to make two decisions sequentially.

1. In the first decision stage, you decide how many tokens to invest into a **Group Account** from your endowment. You can contribute any (whole number) amount between 0 and 20 points.

2. In the second decision stage, you fill out an **"investment table"**. In the investment table you have to indicate for each possible investment of one other group member how many tokens you want to invest in the Group Account. You can condition your investment on the investment of one other group member.

Payment

After all participants have made the two decisions:

One group member will be randomly selected by the system, his/her investment in the first stage will be allocated into the Group Account.
For **other group members**, the decisions in the investment table will be allocated into the Group Account.
Therefore, either of the two decisions you will make could be the one that determines your payoff.

Earnings from Group Account:

Once the investments are determined, total amount of tokens allocated to the Group Account in your group will be **multiplied by 1.6, and then divided equally** between all 4 members in your group.

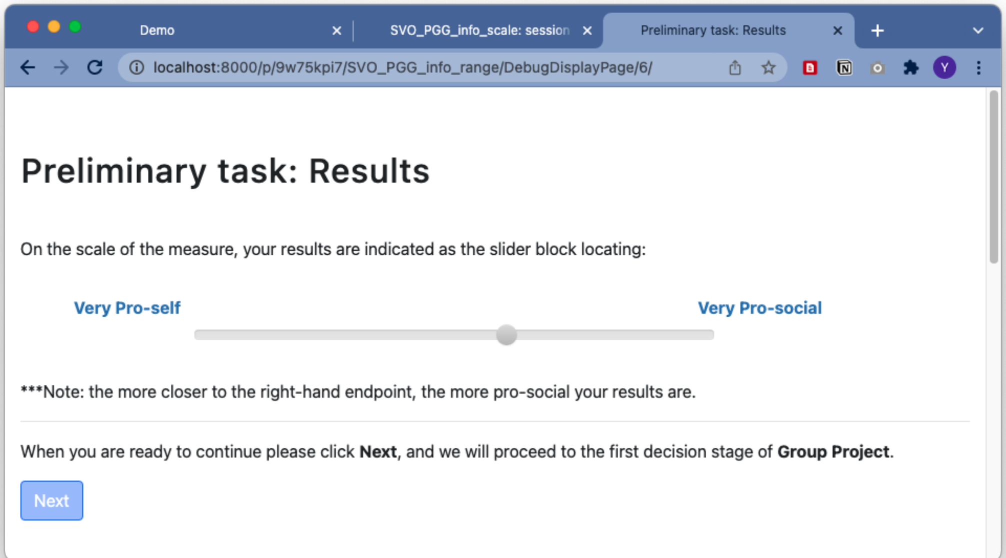
Your earnings will be: Earnings = 20 points - your contribution to Group Account + 1.6 x Total amount in Group Account / 4.

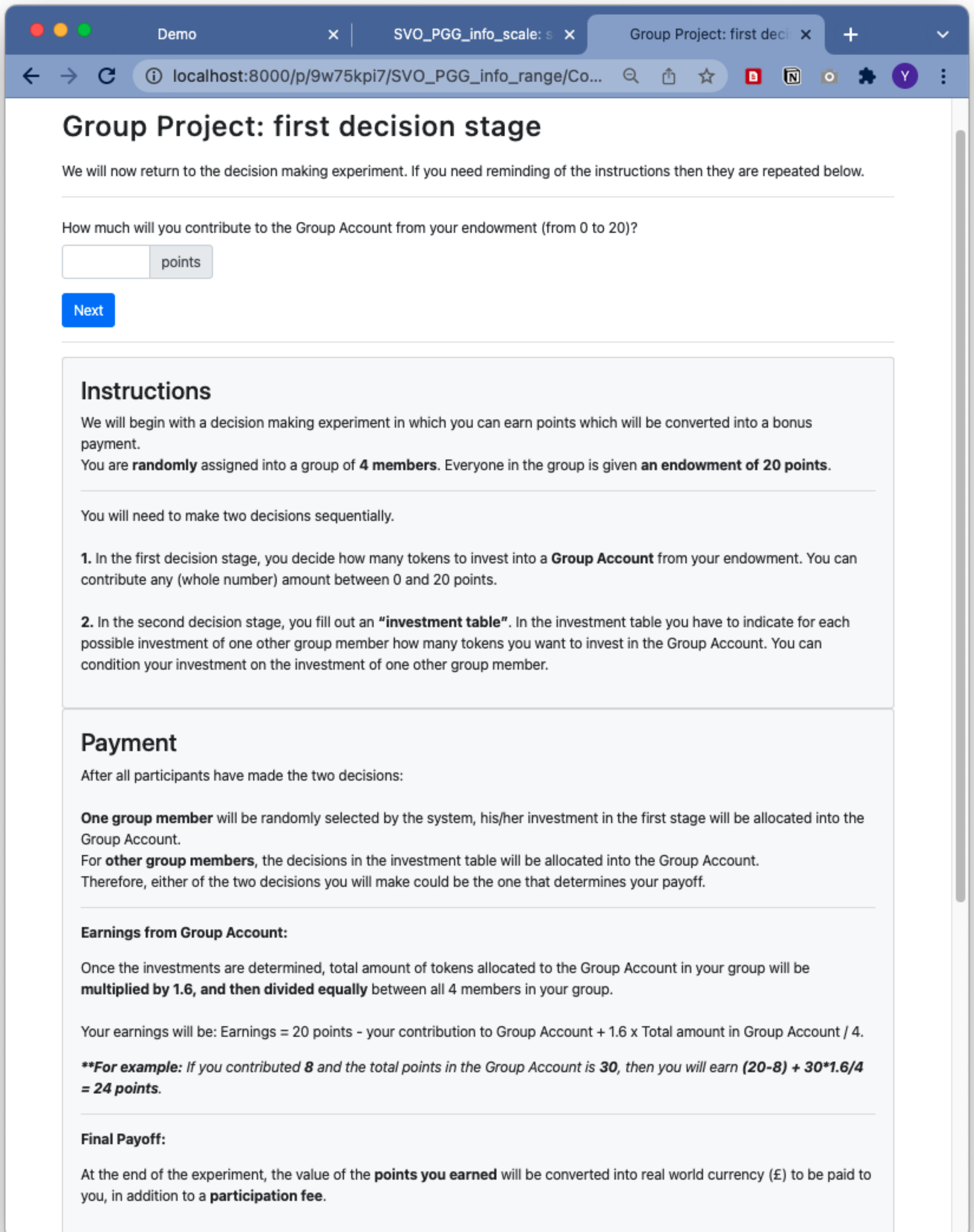
***For example:** *If you contributed **8** and the total points in the Group Account is **30**, then you will earn **(20-8) + 30*1.6/4 = 24 points**.*

Final Payoff:

At the end of the experiment, the value of the **points you earned** will be converted into real world currency (£) to be paid to you, in addition to a **participation fee**.

Next


Demo
SVO_PGG_info_scale: session
Preliminary task: Results
localhost:8000/p/9w75kpi7/SVO_PGG_info_range/DebugDisplayPage/6/
Preliminary task: Results
On the scale of the measure, your results are indicated as the slider block locating:
Very Pro-self
Very Pro-social
***Note: the more closer to the right-hand endpoint, the more pro-social your results are.
When you are ready to continue please click Next, and we will proceed to the first decision stage of Group Project.
Next



Group Project: first decision stage

We will now return to the decision making experiment. If you need reminding of the instructions then they are repeated below.

How much will you contribute to the Group Account from your endowment (from 0 to 20)?

points

Next

Instructions

We will begin with a decision making experiment in which you can earn points which will be converted into a bonus payment.
You are **randomly** assigned into a group of **4 members**. Everyone in the group is given **an endowment of 20 points**.

You will need to make two decisions sequentially.

1. In the first decision stage, you decide how many tokens to invest into a **Group Account** from your endowment. You can contribute any (whole number) amount between 0 and 20 points.

2. In the second decision stage, you fill out an **"investment table"**. In the investment table you have to indicate for each possible investment of one other group member how many tokens you want to invest in the Group Account. You can condition your investment on the investment of one other group member.

Payment

After all participants have made the two decisions:

One group member will be randomly selected by the system, his/her investment in the first stage will be allocated into the Group Account.
For **other group members**, the decisions in the investment table will be allocated into the Group Account.
Therefore, either of the two decisions you will make could be the one that determines your payoff.

Earnings from Group Account:

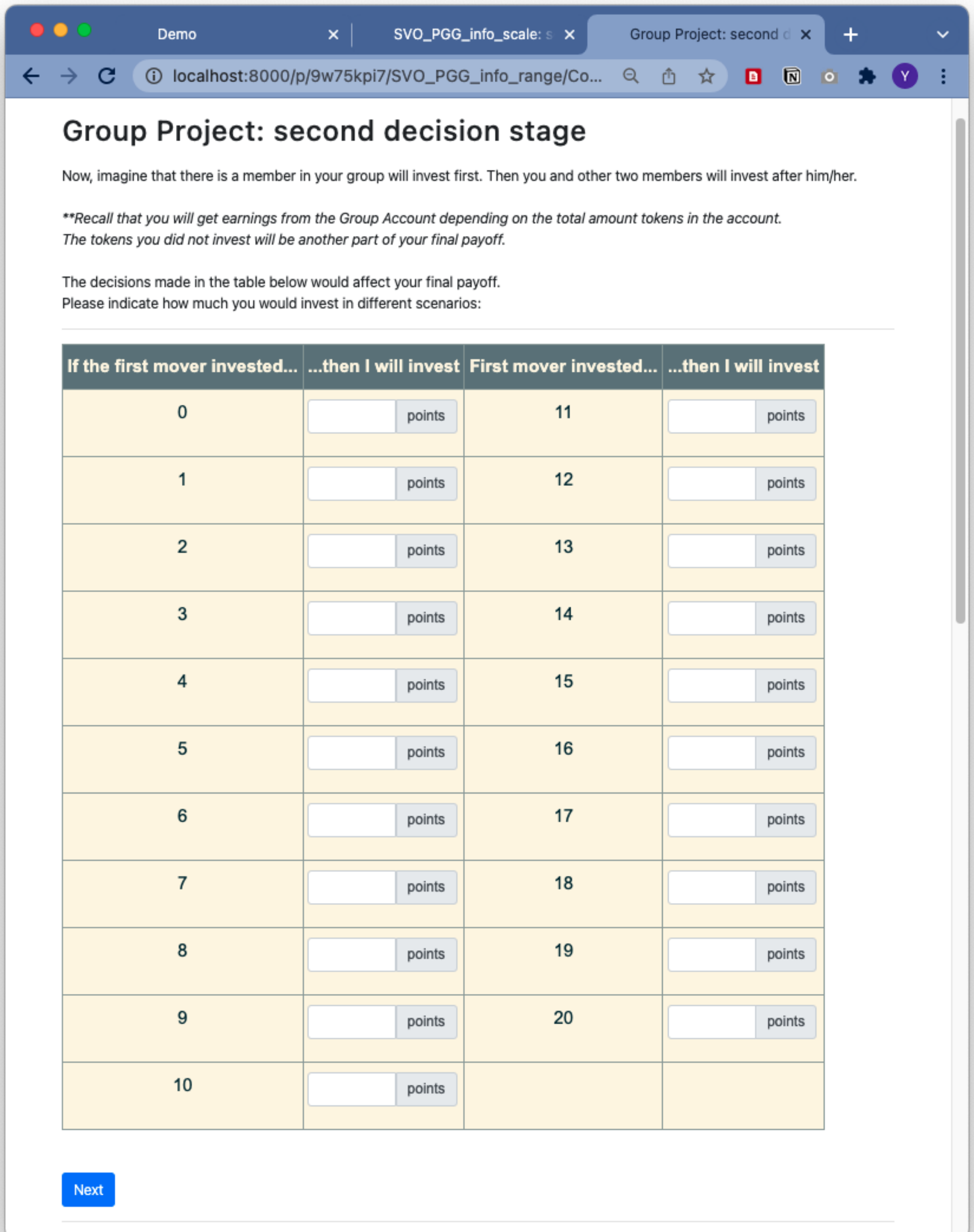
Once the investments are determined, total amount of tokens allocated to the Group Account in your group will be **multiplied by 1.6, and then divided equally** between all 4 members in your group.

Your earnings will be: Earnings = 20 points - your contribution to Group Account + 1.6 x Total amount in Group Account / 4.

****For example:** If you contributed **8** and the total points in the Group Account is **30**, then you will earn **(20-8) + 30*1.6/4 = 24 points**.*

Final Payoff:

At the end of the experiment, the value of the **points you earned** will be converted into real world currency (£) to be paid to you, in addition to a **participation fee**.



Group Project: second decision stage

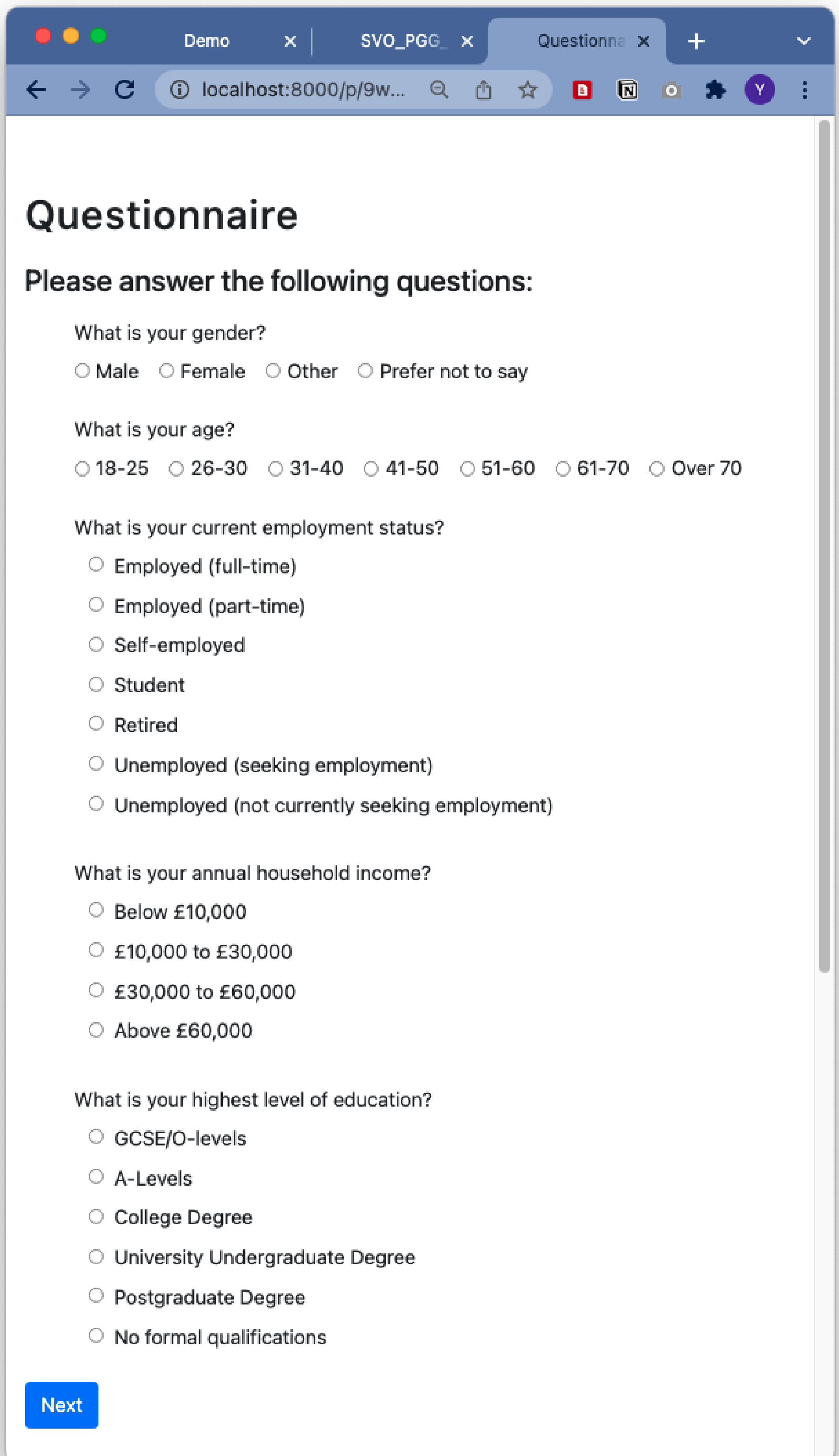
Now, imagine that there is a member in your group will invest first. Then you and other two members will invest after him/her.

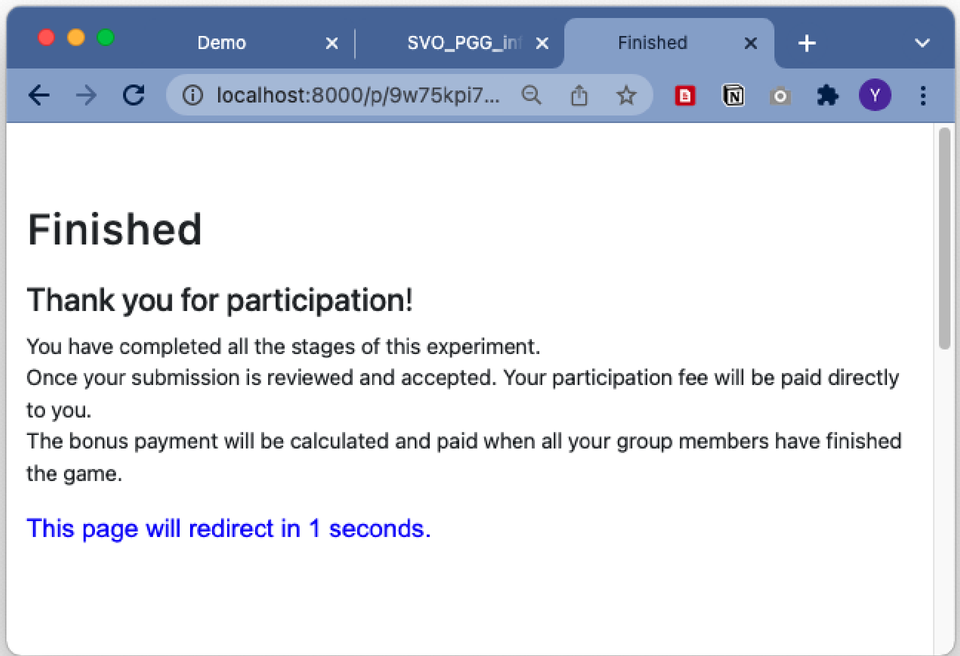
***Recall that you will get earnings from the Group Account depending on the total amount tokens in the account.*
The tokens you did not invest will be another part of your final payoff.

The decisions made in the table below would affect your final payoff.
Please indicate how much you would invest in different scenarios:

If the first mover invested...	...then I will invest	First mover invested...	...then I will invest
0	points	11	points
1	points	12	points
2	points	13	points
3	points	14	points
4	points	15	points
5	points	16	points
6	points	17	points
7	points	18	points
8	points	19	points
9	points	20	points
10	points		

Next


Demo
SVO_PGG_
Questionna
localhost:8000/p/9w...
Questionnaire
Please answer the following questions:
What is your gender?
Male
Female
Other
Prefer not to say
What is your age?
18-25
26-30
31-40
41-50
51-60
61-70
Over 70
What is your current employment status?
Employed (full-time)
Employed (part-time)
Self-employed
Student
Retired
Unemployed (seeking employment)
Unemployed (not currently seeking employment)
What is your annual household income?
Below £10,000
£10,000 to £30,000
£30,000 to £60,000
Above £60,000
What is your highest level of education?
GCSE/O-levels
A-Levels
College Degree
University Undergraduate Degree
Postgraduate Degree
No formal qualifications
Next



A.2.3 Additional Analysis

Table A1: Balance of participants in each treatment: ordered logit regression of treatment on SVO classification and beliefs of SVO

Dependent variable: Treatment *(bl: no information)*	(1)	(2)
SVO type	1.136	
	(0.227)	
Belief of SVO type		0.943
		(0.193)
Gender	0.618**	0.620**
	(0.135)	(0.137)
Age	1.150*	1.156*
	(0.0932)	(0.0933)
Employment status	1.008	1.009
	(0.0692)	(0.0698)
Income	0.968	0.972
	(0.0399)	(0.0401)
Education	1.032	1.035
	(0.0546)	(0.0550)
Observations	342	342
Pseudo R^2	0.0135	0.0131
Wald χ^2	8.793	8.817
Log-likelihood	-369.5	-369.7

Note: Odds ratios from ordered logistic regressions are reported. Dependent variables is a categorical variable with baseline being the no information treatment, 1 and 2 being the Binary and Scale treatment. Overall, the distribution of participants in each treatment is balanced, especially for the key experimental variable (SVO type). Robust standard error in the parenthesis. *, **, and ***indicate significance at the 10%, 5%, and 1% levels, respectively.

A.3 Chapter Four

A.3.1 Experiment Instructions

[Same as Chapter Three, except Group Project: first decision stage]

Leader choice

We now want you to imagine that you are the **leader** in the group. You need

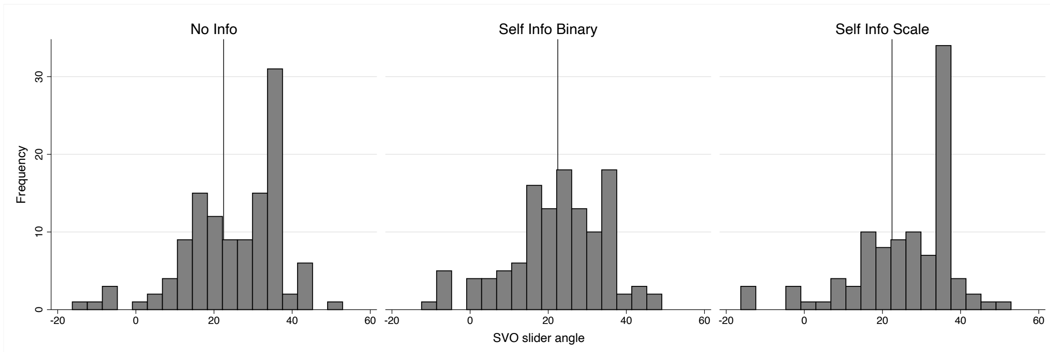


Figure A1: Distribution of participants by SVO slider angle and treatment

to choose a contribution which you would like to invest in the group project. You also have the option of telling followers in the group your SVO type.

Please choose your action below:

(Recall that your SVO type is: [player SVO classification]

[Treatment Cheap-talk]

Q1: I would like to tell them that...

[Participants choose 'Your leader is pro-social' or 'Your leader is pro-self' or 'You will not know the type of the leader']

[Treatment Hide]

Q1: Would you like to tell your group members your Social Value Orientation classification?

[Participants choose 'Yes' or 'No']

[Treatment Truth]

We now want you to imagine that you are the **leader** in the group. You need to choose a contribution which you would like to invest to the group project.

Followers will see your contribution. They may also be told your SVO type.

Specifically, with **70%** probability the followers will be told your type is [player SVO classification]. With **30%** probability your type will not be revealed to followers.

The computer will randomly determine whether your type is revealed or not. This decision will be showed in next page.

[Both treatments]

Q: How much would you like to contribute as a Leader in the group? (from 0 to 5)

[Participants enter a number between 0 to 20]

Leader choice: Revelation Decision

[Only Treatment Truth]

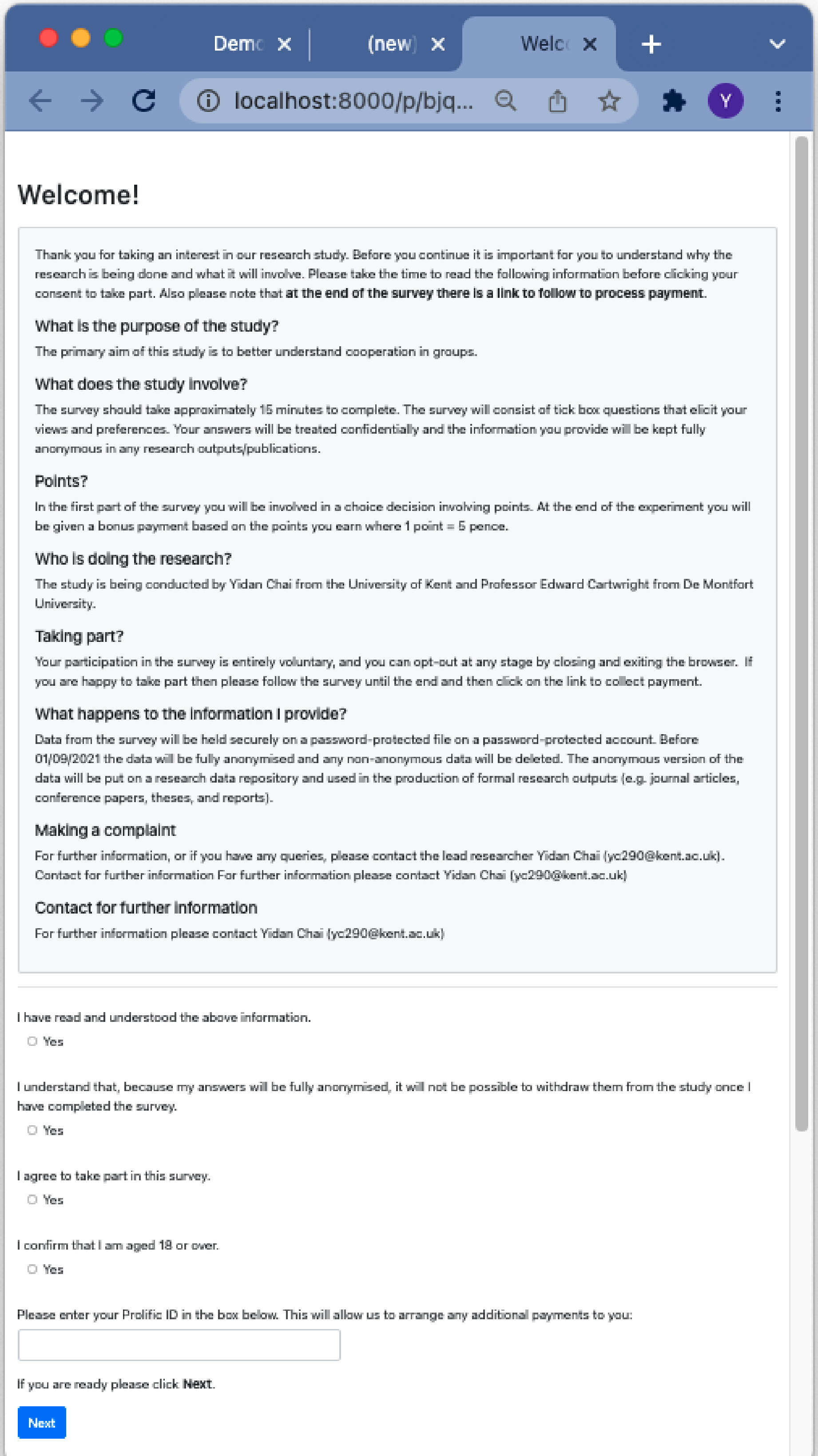
The computer has randomly determined whether your type is revealed or not. The revelation decision is:

[Reveal. Therefore, your members will be told that your type is [player SVO classification].]

[Not Reveal. Therefore, your type will not be informed.]

A.3.2 Experiment Main Pages Screenshots

Here we only presented main pages which are different from experiment 1. Therefore, pages, such as SVO instructions, test, beliefs, results and questionnaire are not included here. Those pages are the same - in terms of structures and order - as what we did in experiment 1.



Welcome!

Thank you for taking an interest in our research study. Before you continue it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information before clicking your consent to take part. Also please note that **at the end of the survey there is a link to follow to process payment**.

What is the purpose of the study?

The primary aim of this study is to better understand cooperation in groups.

What does the study involve?

The survey should take approximately 15 minutes to complete. The survey will consist of tick box questions that elicit your views and preferences. Your answers will be treated confidentially and the information you provide will be kept fully anonymous in any research outputs/publications.

Points?

In the first part of the survey you will be involved in a choice decision involving points. At the end of the experiment you will be given a bonus payment based on the points you earn where 1 point = 5 pence.

Who is doing the research?

The study is being conducted by Yidan Chai from the University of Kent and Professor Edward Cartwright from De Montfort University.

Taking part?

Your participation in the survey is entirely voluntary, and you can opt-out at any stage by closing and exiting the browser. If you are happy to take part then please follow the survey until the end and then click on the link to collect payment.

What happens to the information I provide?

Data from the survey will be held securely on a password-protected file on a password-protected account. Before 01/09/2021 the data will be fully anonymised and any non-anonymous data will be deleted. The anonymous version of the data will be put on a research data repository and used in the production of formal research outputs (e.g. journal articles, conference papers, theses, and reports).

Making a complaint

For further information, or if you have any queries, please contact the lead researcher Yidan Chai (yc290@kent.ac.uk). Contact for further information For further information please contact Yidan Chai (yc290@kent.ac.uk)

Contact for further information

For further information please contact Yidan Chai (yc290@kent.ac.uk)

I have read and understood the above information.

○ Yes

I understand that, because my answers will be fully anonymised, it will not be possible to withdraw them from the study once I have completed the survey.

○ Yes

I agree to take part in this survey.

○ Yes

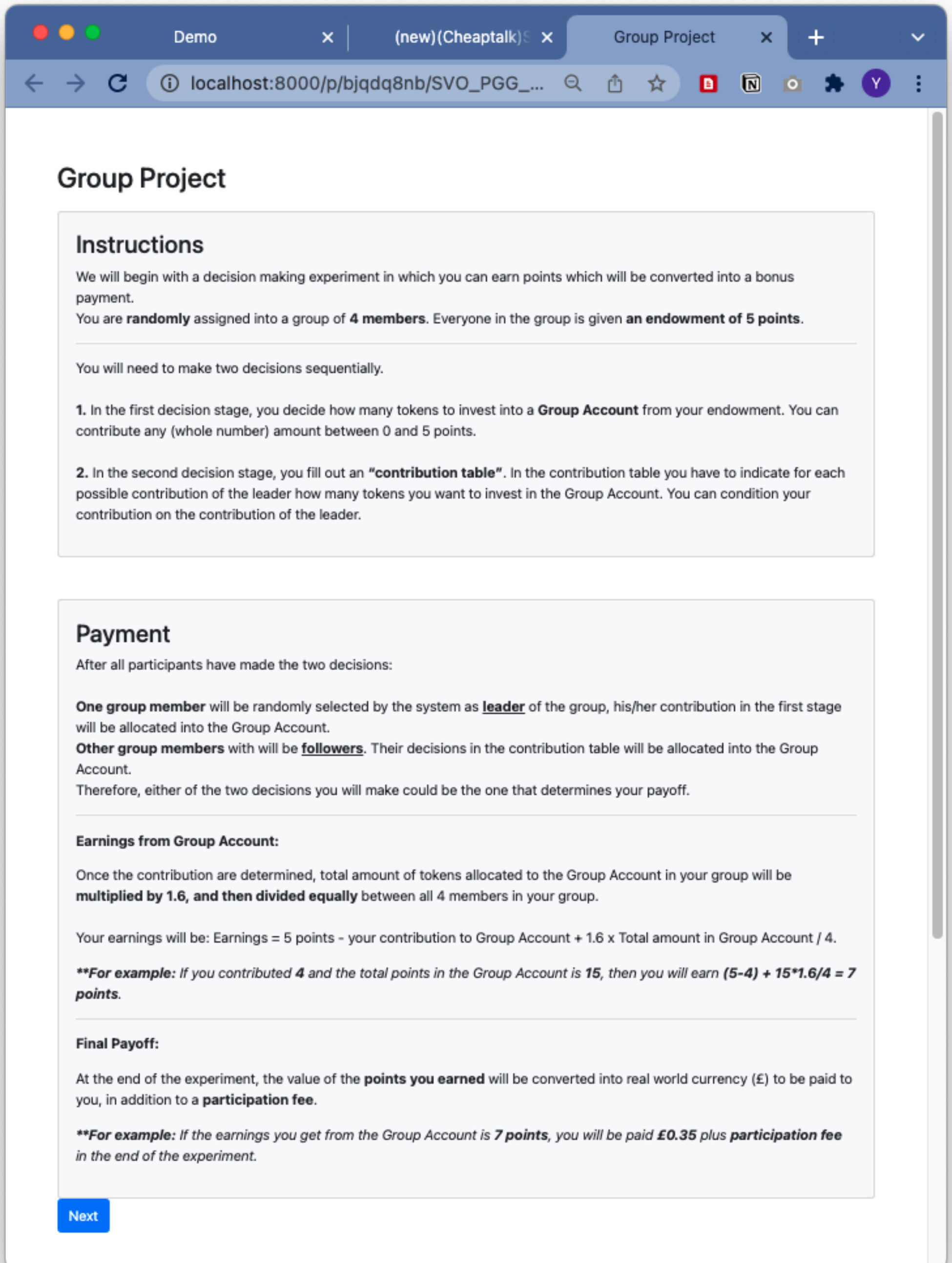
I confirm that I am aged 18 or over.

○ Yes

Please enter your Prolific ID in the box below. This will allow us to arrange any additional payments to you:

If you are ready please click **Next**.

Next



Group Project

Instructions

We will begin with a decision making experiment in which you can earn points which will be converted into a bonus payment.
You are **randomly** assigned into a group of **4 members**. Everyone in the group is given **an endowment of 5 points**.

You will need to make two decisions sequentially.

1. In the first decision stage, you decide how many tokens to invest into a **Group Account** from your endowment. You can contribute any (whole number) amount between 0 and 5 points.

2. In the second decision stage, you fill out an **"contribution table"**. In the contribution table you have to indicate for each possible contribution of the leader how many tokens you want to invest in the Group Account. You can condition your contribution on the contribution of the leader.

Payment

After all participants have made the two decisions:

One group member will be randomly selected by the system as **leader** of the group, his/her contribution in the first stage will be allocated into the Group Account.
Other group members with will be **followers**. Their decisions in the contribution table will be allocated into the Group Account.
Therefore, either of the two decisions you will make could be the one that determines your payoff.

Earnings from Group Account:

Once the contribution are determined, total amount of tokens allocated to the Group Account in your group will be **multiplied by 1.6, and then divided equally** between all 4 members in your group.

Your earnings will be: Earnings = 5 points - your contribution to Group Account + 1.6 x Total amount in Group Account / 4.

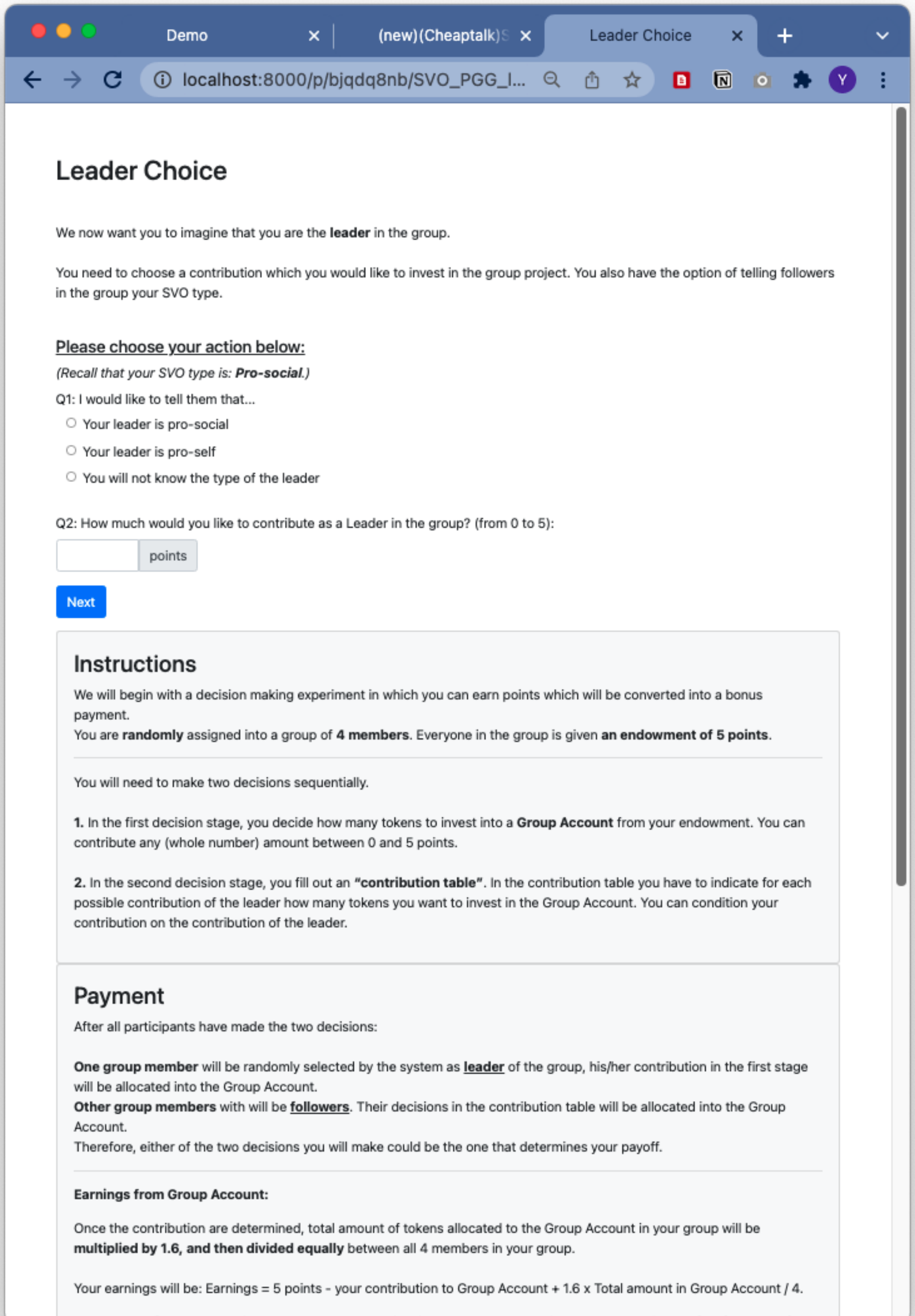
*****For example:** If you contributed **4** and the total points in the Group Account is **15**, then you will earn **(5-4) + 15*1.6/4 = 7 points**.*

Final Payoff:

At the end of the experiment, the value of the **points you earned** will be converted into real world currency (£) to be paid to you, in addition to a **participation fee**.

*****For example:** If the earnings you get from the Group Account is **7 points**, you will be paid **£0.35** plus **participation fee** in the end of the experiment.*

Next



Demo | (new)(Cheaptalk)S | Leader Choice

localhost:8000/p/bjqdq8nb/SVO_PGG_l...

Leader Choice

We now want you to imagine that you are the **leader** in the group.

You need to choose a contribution which you would like to invest in the group project. You also have the option of telling followers in the group your SVO type.

Please choose your action below:

(Recall that your SVO type is: ***Pro-social.****)*

Q1: I would like to tell them that...

- Your leader is pro-social
- Your leader is pro-self
- You will not know the type of the leader

Q2: How much would you like to contribute as a Leader in the group? (from 0 to 5):

points

Next

Instructions

We will begin with a decision making experiment in which you can earn points which will be converted into a bonus payment.
You are **randomly** assigned into a group of **4 members**. Everyone in the group is given **an endowment of 5 points**.

You will need to make two decisions sequentially.

1. In the first decision stage, you decide how many tokens to invest into a **Group Account** from your endowment. You can contribute any (whole number) amount between 0 and 5 points.

2. In the second decision stage, you fill out an **"contribution table"**. In the contribution table you have to indicate for each possible contribution of the leader how many tokens you want to invest in the Group Account. You can condition your contribution on the contribution of the leader.

Payment

After all participants have made the two decisions:

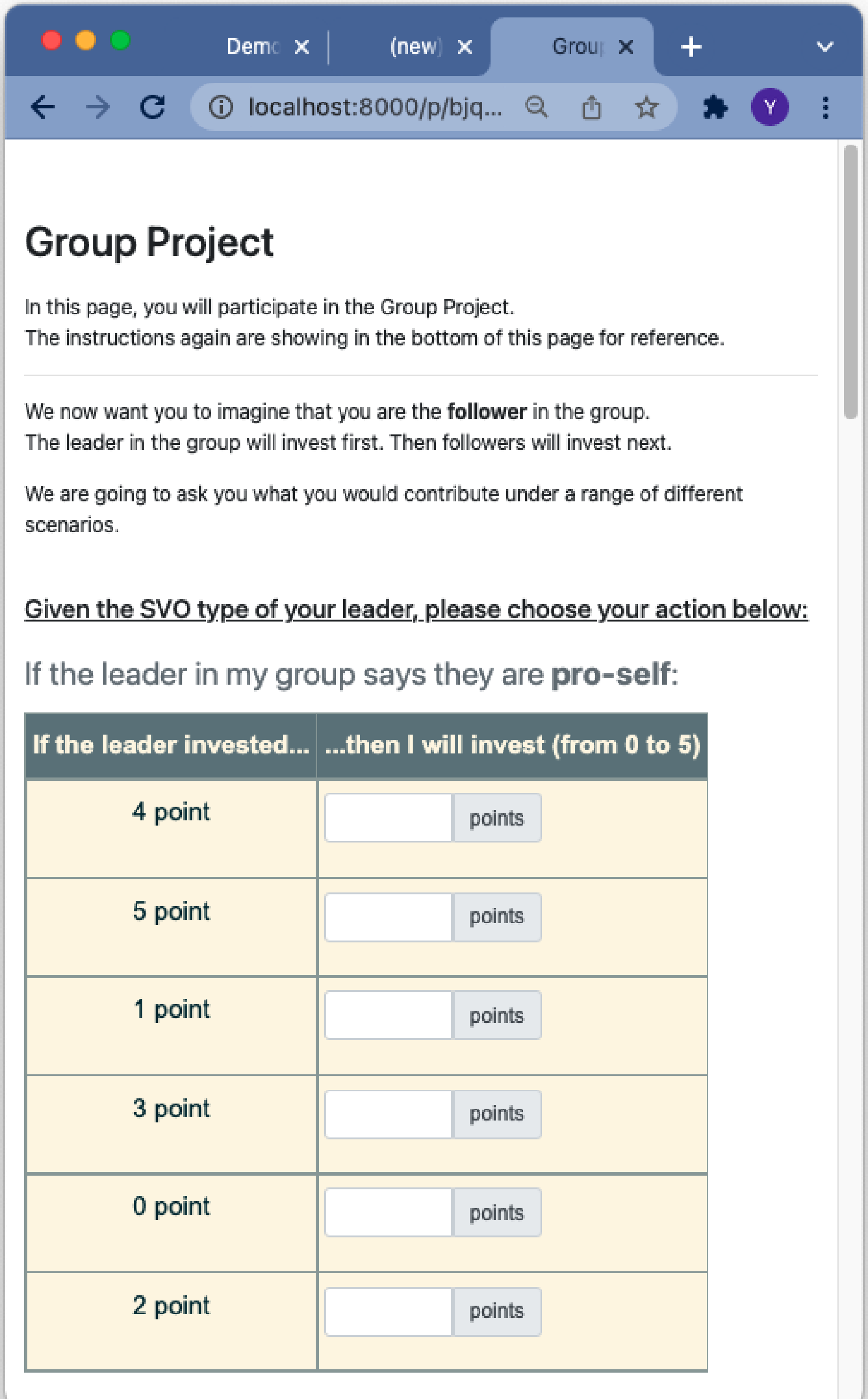
One group member will be randomly selected by the system as **leader** of the group, his/her contribution in the first stage will be allocated into the Group Account.
Other group members with will be **followers**. Their decisions in the contribution table will be allocated into the Group Account.
Therefore, either of the two decisions you will make could be the one that determines your payoff.

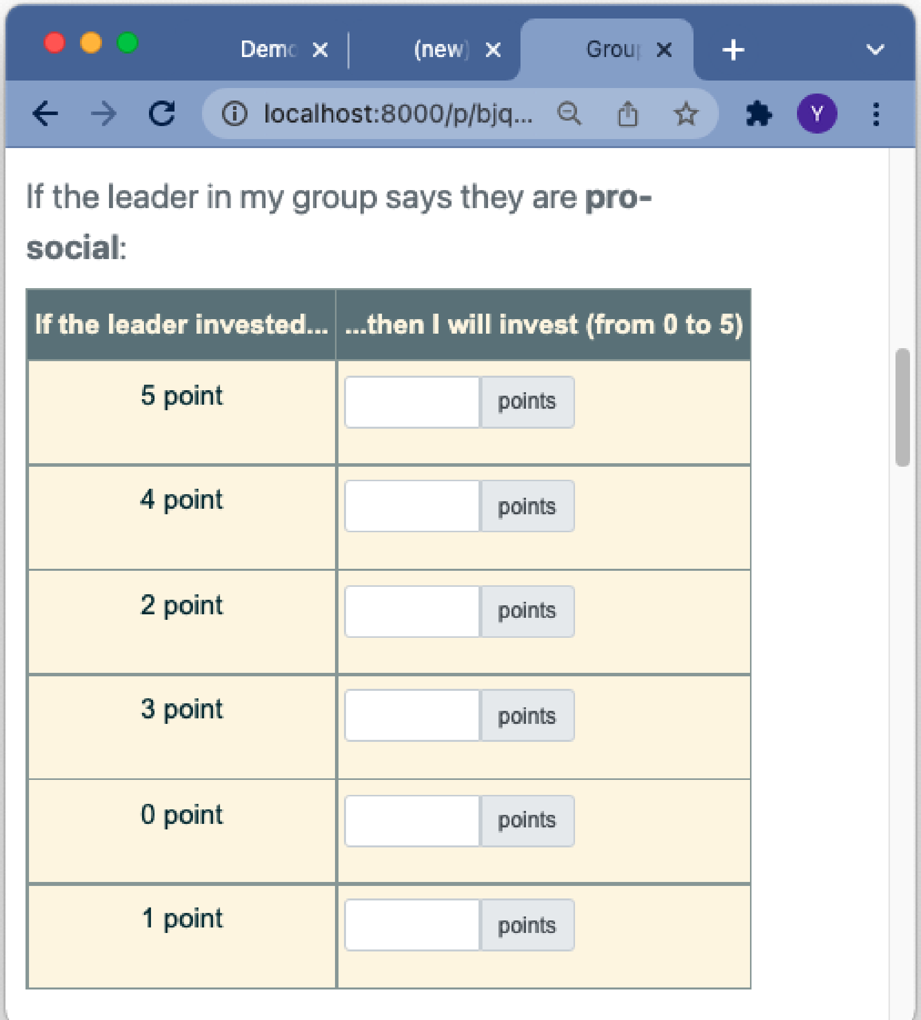
Earnings from Group Account:

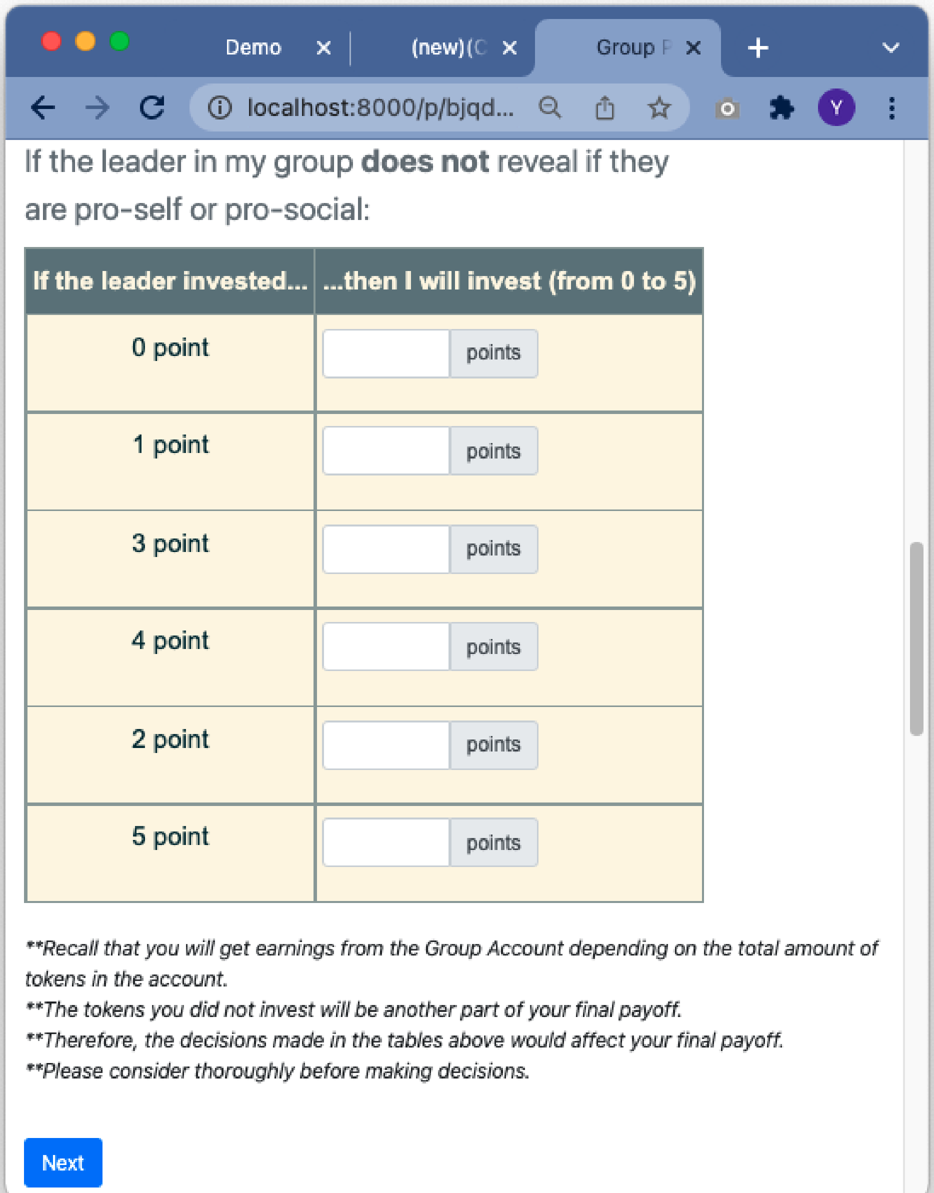
Once the contribution are determined, total amount of tokens allocated to the Group Account in your group will be **multiplied by 1.6, and then divided equally** between all 4 members in your group.

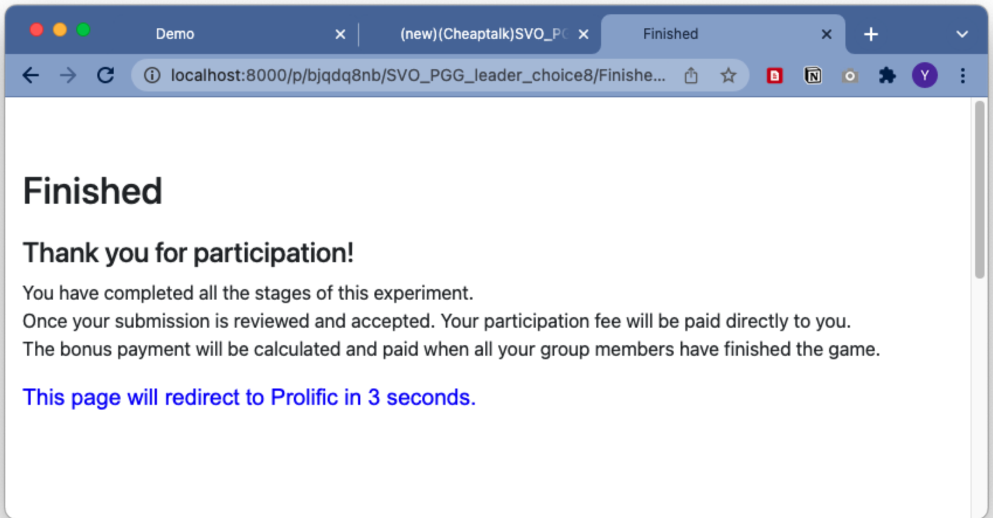
Your earnings will be: Earnings = 5 points - your contribution to Group Account + 1.6 x Total amount in Group Account / 4.

****For example:** *If you contributed* ***4*** *and the total points in the Group Account is* ***15****, then you will earn* ***(5-4) + 15*1.6/4 = 7***


Demo
(new)
Group
localhost:8000/p/bjq...
Y
Group Project
In this page, you will participate in the Group Project.
The instructions again are showing in the bottom of this page for reference.
We now want you to imagine that you are the follower in the group.
The leader in the group will invest first. Then followers will invest next.
We are going to ask you what you would contribute under a range of different scenarios.
Given the SVO type of your leader, please choose your action below:
If the leader in my group says they are pro-self:
If the leader invested...
...then I will invest (from 0 to 5)
4 point
points
5 point
points
1 point
points
3 point
points
0 point
points
2 point
points


Demo
(new)
Group
localhost:8000/p/bjq...
Y
If the leader in my group says they are **pro-social**:
If the leader invested...
...then I will invest (from 0 to 5)
5 point
points
4 point
points
2 point
points
3 point
points
0 point
points
1 point
points


Demo
(new)(C
Group P
localhost:8000/p/bjqd...
If the leader in my group **does not** reveal if they are pro-self or pro-social:
If the leader invested...
...then I will invest (from 0 to 5)
0 point
points
1 point
points
3 point
points
4 point
points
2 point
points
5 point
points
**Recall that you will get earnings from the Group Account depending on the total amount of tokens in the account.
**The tokens you did not invest will be another part of your final payoff.
**Therefore, the decisions made in the tables above would affect your final payoff.
**Please consider thoroughly before making decisions.
Next



A.3.3 Python Script for Simulations

Listing A.1: Simulation Codes

```
import csv
import pandas as pd
import random

from itertools import combinations

filename = "trt3.csv"

csv_reader = pd.read_csv(filename)

with open('sim-truth.csv', mode='w') as sim_file:
    sim_writer = csv.writer(sim_file, delimiter=',', quotechar='"',
    quoting=csv.QUOTE_MINIMAL)
```

```
# FIRST loop
combination_uids = list(combinations(csv_reader.uid,4))
combination_count = len(combination_uids)

for k in range(combination_count):
    print(k)
    one_member_uid = combination_uids[k][0]
    group = csv_reader[csv_reader.uid == one_member_uid]
    for j in range(1, 4):
        group = group.append(csv_reader.loc[csv_reader['uid'] ==
        combination_uids[k][j]])


    # SECOND loop: inside those 4 people

    for n in range(4):

        leader = group.iloc[n]
        leader_uid = leader['uid']
        lucc = leader.playercontribution_l

        # get group total contribution: leader ucc + sum (followers cc)
        fcc = 0
        temp1 = 'playerlprosocial_' + str(lucc)
        temp2 = 'playerlproself_' + str(lucc)
        temp3 = 'playerlunknown_' + str(lucc)
        if leader.message == 'Sc':
            temp = temp1
        elif leader.message == 'Sf':
```

```
    temp = temp2
else:
    temp = temp3
group_sim = group.loc[:, group.columns.str.endswith(temp)]
for i in group_sim.itertuples():
    # print(i)
    if not (i.Index == leader_uid -1):
        fcc += group_sim.at[i.Index,temp]
        # print('follower cc:',fcc)
tt = lucc + fcc

if leader.SVO_type == 'Pro-social':
    leader_sc = 1
else:
    leader_sc = 0

follower_sc_count = 0

for i in group.itertuples():
    if not (i.uid == leader_uid):
        if i.SVO_type == 'Pro-social':
            follower_sc_count += 1

sim_writer.writerow([leader_sc, follower_sc_count, tt, lucc,
leader.message])
```

Bibliography

Kurt A Ackermann and Ryan O Murphy. Explaining cooperative behavior in public goods games: How preferences and beliefs affect contribution levels. *Games*, 10 (1):15, 2019.

David Amdur and Ethan Schmick. Does the direct-response method induce guilt aversion in a trust game? *Available at SSRN 2112715*, 2012.

Christopher M. Anderson and Louis Putterman. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, 54(1):1–24, jan 2006. ISSN 08998256. doi: 10.1016/j.geb.2004.08.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0899825604001654.

Michael L Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484): 1481–1495, 2008.

James Andreoni. Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics*, 37(3):291–304, dec 1988. ISSN 00472727.

doi: 10.1016/0047-2727(88)90043-6. URL https://linkinghub.elsevier.com/retrieve/pii/0047272788900436.

James Andreoni. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401):464, jun 1990. ISSN 00130133. doi: 10.2307/2234133. URL https://academic.oup.com/ej/article/100/401/464-477/5190270.

James Andreoni and Rachel Croson. Chapter 82 Partners versus Strangers: Random Rematching in Public Goods Experiments. *Handbook of Experimental Economics Results*, 1(C):776–783, 2008. ISSN 15740722. doi: 10.1016/S1574-0722(07)00082-0.

James Andreoni and John Miller. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2):737–753, mar 2002. ISSN 0012-9682. doi: 10.1111/1468-0262.00302. URL http://www.jstor.org/stable/2692289http://doi.wiley.com/10.1111/1468-0262.00302.

James Andreoni and Ragan Petrie. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of public Economics*, 88:1605–1623, 2004.

Emrah Arbak and Marie-Claire Villeval. Voluntary leadership: motivation and influence. *Social Choice and Welfare*, 40(3):635–662, 2013.

Jalal Eddine Bahbouhi and Najem Moussa. A graph-based model for public goods with leaderships. *Applied Mathematics and Computation*, 349:53–61, 2019.

Daniel Balliet, Craig Parks, and Jeff Joireman. Social Value Orientation and Cooperation in Social Dilemmas: A Meta-Analysis. *Group Processes & Intergroup Re-*

lations, 12(4):533–547, jul 2009. ISSN 1368-4302. doi: 10.1177/1368430209105040. URL http://journals.sagepub.com/doi/10.1177/1368430209105040.

Pierpaolo Battigalli and Martin Dufwenberg. Guilt in games. *American Economic Review*, 97(2):170–176, 2007.

Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144:1–35, 2009.

Pierpaolo Battigalli and Martin Dufwenberg. Psychological game theory. Technical report, 2019.

Roy F Baumeister, Arlene M Stillwell, and Todd F Heatherton. Guilt: an interpersonal approach. *Psychological bulletin*, 115(2):243, 1994.

R Meredith Belbin. Management teams: Why they succeed or fail. *Human Resource Management International Digest*, 2011.

R Meredith Belbin. *Team roles at work*. Routledge, 2012.

Charles Bellemare, Alexander Sebald, and Martin Strobel. Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453, 2011.

Roland Bénabou and Jean Tirole. Incentives and Prosocial Behavior. *American Economic Review*, 96(5):1652–1678, nov 2006. ISSN 0002-8282. doi: 10.1257/aer.96.5.1652. URL http://ideas.repec.org/a/aea/aecrev/v96y2006i5p1652-1678.htmlhttp://pubs.aeaweb.org/doi/10.1257/aer.96.5.1652.

Kornpob Bhirombhakdi and Tanapong Potipiti. Performance of a reciprocity model in predicting a positive reciprocity decision. *SOUTHEAST ASIAN JOURNAL OF ECONOMICS*, pages 77–93, 2011.

Cristina Bicchieri. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press, 2016.

Cristina Bicchieri, Ryan Muldoon, Alessandro Sontuoso, et al. Social norms, 2018.

Maria Bigoni and Sigrid Suetens. Feedback and dynamics in public good experiments. *Journal of Economic Behavior and Organization*, 82:86–95, 2012.

Mariana Blanco, Dirk Engelmann, Alexander K Koch, and Hans-Theo Normann. Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, 13(4):412–438, 2010.

Mariana Blanco, Dirk Engelmann, and Hans Theo Normann. A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2): 321–338, 2011. ISSN 08998256. doi: 10.1016/j.geb.2010.09.008. URL http://dx.doi.org/10.1016/j.geb.2010.09.008.

Mariana Blanco, Dirk Engelmann, Alexander K Koch, and Hans-Theo Normann. Preferences and beliefs in a sequential social dilemma: a within-subjects analysis. *Games and Economic Behavior*, 87:122–135, 2014.

Sandy Bogaert, Christophe Boone, and Carolyn Declerck. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, 47(3):453–480, sep 2008. ISSN 01446665. doi: 10.1348/014466607X244970. URL http://doi.wiley.com/10.1348/014466607X244970.

Peter Bohm. Estimating demand for public goods: An experiment. *European Economic Review*, 3(2):111–130, jan 1972. ISSN 00142921. doi: 10.1016/0014-2921(72)90001-3. URL https://linkinghub.elsevier.com/retrieve/pii/0014292172900013.

Gary E. Bolton. A Comparative Model of Bargaining : Theory and Evidence. *The American Economic Review*, 81(5):1096–1136, 1991. URL http://www.jstor.org/stable/2006908.

Gary E. Bolton and Axel Ockenfels. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1):166–193, mar 2000. ISSN 0002-8282. doi: 10.1257/aer.90.1.166. URL http://pubs.aeaweb.org/doi/10.1257/aer.90.1.166.

Luke A. Boosey. Conditional cooperation in network public goods experiments. *Journal of Behavioral and Experimental Economics*, 69:108–116, 2017.

Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret L. Satterthwaite, Oded Nov, and Enrico Bertini. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.

Robert Boyd and Peter J. Richerson. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3):337–356, jun 1988. ISSN 00225193. doi: 10.1016/S0022-5193(88)80219-4. URL http://dx.doi.org/10.1016/j.jtbi.2009.01.004https://linkinghub.elsevier.com/retrieve/pii/S0022519388802194.

Robert Boyd, Herbert Gintis, Samuel Bowles, and Peter J Richerson. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100 (6):3531–3535, mar 2003. ISSN 0027-8424. doi: 10.1073/pnas.0630443100. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0630443100.

Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007. ISSN 00220531. doi: 10.1016/j.jet.2006.06.006.

Jordi Brandts, David J. Cooper, and Roberto A. Weber. Legitimacy, communication, and leadership in the turnaround game. *Management Science*, 61:2627–2645, 2015.

Jerry M Burger, Jackeline Sanchez, Jenny E Imberi, and Lucia R Grande. The norm of reciprocity as an internalized social norm: Returning favors even when no one finds out. *Social Influence*, 4(1):11–17, 2009.

Maxwell N Burton-Chellew, Claire El Mouden, and Stuart A West. Conditional cooperation and confusion in public-goods experiments. *Proceedings of the National Academy of Sciences*, 113(5):1291–1296, 2016.

Robert Böhm and Maik MP Theelen. Outcome valence and externality valence framing in public good dilemmas. *Journal of Economic Psychology*, 54:151–163, 2016.

Antoni Calvó-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *Review of Economic Studies*, 76(4):1239–1267, 2009. ISSN 00346527. doi: 10.1111/j.1467-937X.2009.00550.x.

A. Stefano Caria and Marcel Fafchamps. Expectations, network centrality, and public good contributions: Experimental evidence from India. *Journal of Economic Behavior & Organization*, (June), aug 2017. ISSN 01672681. doi: 10.1016/j.jebo.2017.07.024. URL https://linkinghub.elsevier.com/retrieve/pii/S0167268117302020.

Jeffrey P. Carpenter, Peter Hans Matthews, and Okomboli Ong'ong'a. Why punish? Social reciprocity and the enforcement of prosocial norms. *Journal of Evolutionary Economics*, 14(4):407–429, 2004. ISSN 09369937. doi: 10.1007/s00191-004-0212-1.

Jeffrey P. Carpenter, Shachar Kariv, and Andrew Schotter. Network architecture, cooperation and punishment in public good experiments. *Review of Economic Design*, 16:93–118, 2012.

José A. Carrasco, Rodrigo Harrison, and Mauricio Villena. Interdependent preferences and endogenous reciprocity. *Journal of Behavioral and Experimental Economics*, 76:68–75, 2018. ISSN 22148051. doi: 10.1016/j.socec.2018.08.002.

Edward Cartwright. A survey of belief-based guilt aversion in trust and dictator games. *Journal of Economic Behavior & Organization*, 167:430–444, 2019.

Edward Cartwright and Denise Lovett. Conditional cooperation and the marginal per capita return in public good games. *Games*, 5(4):234–256, 2014.

Edward Cartwright and Amrish Patel. Imitation and the incentive to contribute early in a sequential public good game. *Journal of Public Economic Theory*, 12 (4):691–708, 2010.

Edward Cartwright, Joris Gillet, and Mark Van Vugt. Leadership by example in the weak-link game. *Economic Inquiry*, 51(4):2028–2043, 2013.

Alessandra Cassar. Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behavior*, 58:209–230, 2007.

Luke J. Chang, Alec Smith, Martin Dufwenberg, and Alan G. Sanfey. Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, 70: 560–572, 2011.

Gary Charness and Martin Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006.

Gary Charness and Martin Dufwenberg. Bare promises: An experiment. *Economics letters*, 107(2):281–283, 2010.

Gary Charness and Matthew Rabin. Some Simple Tests of Preferences and a New Model. 2000.

Gary Charness and Matthew Rabin. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869, aug 2002. ISSN 0033-5533. doi: 10.1162/003355302760193904. URL http://elsa.berkeley.edu/rabin/index.html.https://academic.oup.com/qje/article-lookup/doi/10.1162/0033.

Ananish Chaudhuri. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83, 2011. ISSN 13864157. doi: 10.1007/s10683-010-9257-1.

Ananish Chaudhuri and Tirnud Paichayontvijit. Conditional Cooperation and Voluntary Contributions to Public Goods. *Economics Bulletin*, 3(8):1–14, 2006. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.498.1666.

Ananish Chaudhuri, Tirnud Paichayontvijit, and Alexander Smith. Belief heterogeneity and contributions decay among conditional cooperators in public goods games. *Journal of Economic Psychology*, 58:15–30, 2017.

Daniel L. Chen, Martin Schonger, and Chris Wickens. otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

Dan S. Chiaburu, Troy A. Smith, Jiexin Wang, and Ryan D. Zimmerman. Relative importance of leader influences for subordinates' proactive behaviors, prosocial behaviors, and task performance. *Journal of Personnel Psychology*, 2014.

Syngjoo Choi, Douglas Gale, Shachar Kariv, and Thomas Palfrey. Network architecture, salience and coordination. *Games and Economic Behavior*, 73 (1):76–90, 2011. ISSN 08998256. doi: 10.1016/j.geb.2011.01.001. URL http://dx.doi.org/10.1016/j.geb.2011.01.001.

James C. Cox, Daniel Friedman, and Steven Gjerstad. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45, apr 2007. ISSN 08998256. doi: 10.1016/j.geb.2006.05.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0899825606000662.

Rachel TA Croson. The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes*, 80(2):118–133, 1999.

Rachel TA Croson. Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of economic behavior & organization*, 41(3):299–314, 2000.

Rachel TA Croson. Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, 45:199–216, 2007.

Robin Cubitt, Simon Gächter, and Simone Quercia. Conditional cooperation and betrayal aversion. *Journal of Economic Behavior & Organization*, 141:110–121, 2017.

Robyn M. Dawes, Jeanne McTavish, and Harriet Shaklee. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, 35(1):1–11, 1977. ISSN 0022-3514. doi: 10.1037/0022-3514.35.1.1. URL http://content.apa.org/journals/psp/35/1/1.

David De Cremer and Eric van Dijk. Reactions to group success and failure as a function of identification level: a test of the goal-transformation hypothesis in social dilemmas. *Journal of Experimental Social Psychology*, 38(5):435–442, sep 2002. ISSN 00221031. doi: 10.1016/S0022-1031(02)00009-4. URL http://www.sciencedirect.com/science/article/pii/S0022103102000094https://linkinghub.elsevier.com/retrieve/pii/S0022103102000094.

David De Cremer and Paul A. M. Van Lange. Why prosocials exhibit greater cooperation than proselfs: the roles of social responsibility and reciprocity. *European Journal of Personality*, 15(S1):S5–S18, nov 2001. ISSN 0890-2070. doi: 10.1002/per.418. URL http://doi.wiley.com/10.1002/per.418.

David De Cremer and Mark Van Vugt. Social identification effects in social dilemmas: A transformation of motives. *European Journal of Social Psychology*, 29:871–893, 1999.

David De Cremer and Mark Van Vugt. Intergroup and intragroup aspects of leadership in social dilemmas: A relational model of cooperation. *Journal of Experimental Social Psychology*, 38(2):126–136, 2002.

David De Cremer, Daan Van Knippenberg, Eric Van Dijk, and Esther Van Leeuwen. Cooperating if one's goals are collective-based: Social identification effects in social dilemmas as a function of goal transformation 1. *Journal of Applied Social Psychology*, 38:1562–1579, 2008.

Carsten K.W. De Dreu and Christopher McCusker. Gain-loss frames and cooperation in two-person social dilemmas: A transformational analysis. *Journal of Personality and Social Psychology*, 1997. ISSN 00223514. doi: 10.1037/0022-3514.72.5.1093.

Angela C. M. de Oliveira, Rachel T. A. Croson, and Catherine Eckel. One bad apple? Heterogeneity and information in public good provision. *Experimental Economics*, 18(1):116–135, mar 2015. ISSN 1386-4157. doi: 10.1007/s10683-014-9412-1. URL http://link.springer.com/10.1007/s10683-014-9412-1.

Geert Dhaene and Jan Bouckaert. Sequential reciprocity in two-player, two-stage games: An experimental analysis. *Games and Economic Behavior*, 70(2):289–303, 2010.

Sanjit Dhami, Mengxing Wei, and Ali al Nowaihi. Public goods games and psychological utility: Theory and evidence. *Journal of Economic Behavior & Organization*, 167:361–390, 2019.

Michalis Drouvelis and Nikolaos Georgantzis. Does revealing personality data affect prosocial behaviour? *Journal of Economic Behavior & Organization*, (xxxx), 2019. ISSN 01672681. doi: 10.1016/j.jebo.2019.02.019. URL https://linkinghub.elsevier.com/retrieve/pii/S0167268119300484.

Martin Dufwenberg and Georg Kirchsteiger. A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298, 2004. ISSN 08998256. doi: 10.1016/j.geb.2003.06.003.

Martin Dufwenberg and Amrish Patel. Reciprocity networks and the participation problem. *Games and Economic Behavior*, 101:260–272, 2017.

Martin Dufwenberg, Simon Gächter, and Heike Hennig-Schmidt. The framing of games and the psychology of play. *Games and Economic Behavior*, 2011. ISSN 08998256. doi: 10.1016/j.geb.2011.02.003.

Catherine C. Eckel, Enrique Fatas, and Rick Wilson. Cooperation and status in organizations. *Journal of Public Economic Theory*, 12:737–762, 2010.

Tony Edwards, Paul Marginson, Paul Edwards, Anthony Ferner, and Olga Tregaskis. *Corporate social responsibility in multinational companies: Management initiatives or negotiated agreements?* Internat. Inst. for Labour Studies, 2007.

Tore Ellingsen, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik. Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107, 2010.

Tore Ellingsen, Magnus Johannesson, Johanna Mollerstrom, and Sara Munkhammar. Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1):117–130, 2012.

Matthew Elliott and Benjamin Golub. A Network Approach to Public Goods. *Journal of Political Economy*, 127(2):730–776, apr 2019. ISSN 0022-3808. doi: 10.

1086/701032. URL https://www.journals.uchicago.edu/doi/pdfplus/10.1086/701032https://www.journals.uchicago.edu/doi/10.1086/701032.

John Emerson, Margaret L. Satterthwaite, Anshul Vikram Pandey, Oded Nov, and Enrico Bertini. *The Challenging Power of Data Visualization for Human Rights Advocacy*. Cambridge University Press, 2016 - in press.

Dirk Engelmann and Martin Strobel. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review*, 94(4):857–869, aug 2004. ISSN 0002-8282. doi: 10.1257/0002828042002741. URL http://pubs.aeaweb.org/doi/abs/10.1257/aer.96.5.1906http://pubs.aeaweb.org/doi/10.1257/0002828042002741.

Mathias Erlei. Heterogeneous social preferences. *Journal of Economic Behavior & Organization*, 65(3-4):436–457, mar 2008. ISSN 01672681. doi: 10.1016/j.jebo.2006.02.006. URL https://linkinghub.elsevier.com/retrieve/pii/S016726810600206X.

Armin Falk and Urs Fischbacher. Distributional Consequences and Intentions in a Model of Reciprocity. *Annales d'Économie et de Statistique*, (63/64):111, 2001. ISSN 0769489X. doi: 10.2307/20076298. URL https://www.jstor.org/stable/10.2307/20076298.

Armin Falk and Urs Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315, 2006. ISSN 08998256. doi: 10.1016/j.geb.2005.03.001.

Armin Falk, Ernst Fehr, and Urs Fischbacher. Informal Sanctions. *Ssrn*, (59), 2000. doi: 10.2139/ssrn.245568.

Armin Falk, Ernst Fehr, and Urs Fischbacher. On the Nature of Fair Behavior. *Economic Inquiry*, 41(1):20–26, jan 2003. ISSN 00952583. doi: 10.1093/ei/41.1.20. URL http://doi.wiley.com/10.1093/ei/41.1.20.

Armin Falk, Ernst Fehr, and Urs Fischbacher. Driving Forces Behind Informal Sanctions. *Econometrica*, 73(6):2017–2030, nov 2005. ISSN 0012-9682. doi: 10.1111/j.1468-0262.2005.00644.x. URL http://www.jstor.org/stable/3598757{%}5Cnhttp://about.jstor.org/termshttp://doi.wiley.com/10.1111/j.1468-0262.2005.00644.x.

Armin Falk, Ernst Fehr, and Urs Fischbacher. Testing theories of fairness-Intentions matter. *Games and Economic Behavior*, 62(1):287–303, 2008. ISSN 08998256. doi: 10.1016/j.geb.2007.06.001.

Armin Falk, Urs Fischbacher, and Simon Gachter. Living in two neighborhoods—social interaction effects in the laboratory. *Economic Inquiry*, 51:563–578, 2013.

Ernst Fehr. Theories of Fairness and Reciprocity: Evidence and Economic Applications. In *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress Vol I*, pages 208–257. 2008. doi: 10.1017/ccol0521818729.007.

Ernst Fehr and Urs Fischbacher. Why social preferences matter - The impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478):C1–C33, mar 2002. ISSN 00130133. doi: 10.1111/1468-0297.00027. URL https://academic.oup.com/ej/article/112/478/C1/5085630.

Ernst Fehr and Urs Fischbacher. The nature of human altruism. *Nature*, 425

(6960):785–791, oct 2003. ISSN 0028-0836. doi: 10.1038/nature02043. URL http://www.nature.com/articles/nature02043.

Ernst Fehr and Urs Fischbacher. Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2):63–87, mar 2004. ISSN 10905138. doi: 10.1016/S1090-5138(04)00005-4. URL https://linkinghub.elsevier.com/retrieve/pii/S1090513804000054.

Ernst Fehr and Simon Gächter. Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14(3):159–182, aug 2000a. ISSN 0895-3309. doi: 10.1257/jep.14.3.159. URL http://pubs.aeaweb.org/doi/10.1257/jep.14.3.159.

Ernst Fehr and Simon Gächter. Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4):980–994, sep 2000b. ISSN 0002-8282. doi: 10.1257/aer.90.4.980. URL http://www.ncbi.nlm.nih.gov/pubmed/24473658http://pubs.aeaweb.org/doi/10.1257/aer.90.4.980.

Ernst Fehr and Simon Gächter. Erratum: Altruistic punishment in humans. *Nature*, 415(6868):137–140, jan 2002. ISSN 0028-0836. doi: 10.1038/415137a. URL http://www.nature.com/articles/415137a.

Ernst Fehr and Klaus M. Schmidt. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, aug 1999. ISSN 0033-5533. doi: 10.1162/003355399556151. URL https://academic.oup.com/qje/article-lookup/doi/10.1162/003355399556151.

Ernst Fehr and Klaus M. Schmidt. Chapter 8 The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories. In *Handbook of* *the Economics of Giving, Altruism and Reciprocity*, volume 1, pages 615–691. 2006. ISBN 9780444506979. doi: 10.1016/S1574-0714(06)01008-6. URL https://linkinghub.elsevier.com/retrieve/pii/S1574071406010086.

Ernst Fehr and Ivo Schurtenberger. Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458–468, 2018.

Ernst Fehr, Simon Gächter, and Georg Kirchsteiger. Reciprocal fairness and noncompensating wage differentials. *Journal of Institutional and Theoretical Economics*, 1996. ISSN 0932-4569.

Cristian Felix, Anshul Vikram Pandey, Enrico Bertini, Charles Ornstein, and Scott Klein. Revex: Visual investigative journalism with a million healthcare reviews. *Computation + Journalism Symposium*, 2015.

Cristian Felix, Anshul Vikram Pandey, and Enrico Bertini. Texttile: An interactive visualization tool for seamless exploratory analysis of structured data and unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, Jan 2017.

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

Urs Fischbacher and Simon Gächter. Heterogeneous social preferences and the dynamics of free riding in public goods. 2006.

Urs Fischbacher and Simon Gächter. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review*, 100

(1):541–556, mar 2010. ISSN 0002-8282. doi: 10.1257/aer.100.1.541. URL http://pubs.aeaweb.org/doi/10.1257/aer.100.1.541.

Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71 (3):397–404, jun 2001. ISSN 01651765. doi: 10.1016/S0165-1765(01)00394-9. URL http://dx.doi.org/10.1016/j.joep.2011.05.007{%}0Apapers://13507515-a992-4cea-b488-dec1e41c1983/Paper/p4802http://linkinghub.elsevier.com/retrieve/pii/S0165176501003949.

Simon Gächter. Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. 2006. URL https://www.econstor.eu/bitstream/10419/67977/1/512169225.pdf.

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

Simon Gächter and Elke Renner. Leaders as role models and 'belief managers' in social dilemmas. *Journal of Economic Behavior & Organization*, 154:321–334, 2018.

Simon Gächter, Daniele Nosenzo, Elke Renner, and Martin Sefton. Sequential versus simultaneous contributions to public goods: Experimental evidence. 2009.

Simon Gächter, Daniele Nosenzo, Elke Renner, and Martin Sefton. Who makes a good leader? cooperativeness, optimism, and leading-by-example. *Economic Inquiry*, 50(4):953–967, 2012.

Tommy Gärling, Satoshi Fujii, Anita Gärling, and Cecilia Jakobsson. Moderating effects of social value orientation on determinants of proenvironmental behavior intention. *Journal of Environmental Psychology*, 23(1):1–9, mar 2003. ISSN 02724944. doi: 10.1016/S0272-4944(02)00081-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0272494402000816.

Sergey Gavrilets and Peter J Richerson. Collective action and the evolution of social norm internalization. *Proceedings of the National Academy of Sciences*, 114(23):6068–6073, 2017.

John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and economic Behavior*, 1(1):60–79, 1989.

Joris Gillet, Edward Cartwright, and Mark Van Vugt. Selfish or servant leadership? evolutionary predictions on leadership personalities in coordination games. *Personality and Individual Differences*, 51:231–236, 2011.

Herbert Gintis et al. The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions. *Journal of economic behavior & organization*, 53(1):57–67, 2004.

Andreas Glöckner, Bernd Irlenbusch, Sebastian Kube, Andreas Nicklisch, and Hans-Theo Normann. Leading with (out) sacrifice? a public-goods experiment with a privileged player. *Economic Inquiry*, 49(2):591–597, 2011.

Timo Goeschl and Johannes Lohse. Cooperation in public good games. calculated or confused? *European Economic Review*, 107:185–203, 2018.

Werner Güth, M Vittoria Levati, Matthias Sutter, and Eline Van Der Heijden. Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042, 2007.

Simon Gächter, Felix Kölle, and Simone Quercia. Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, 1:650–656, 2017.

Stefan D Haigner and Florian Wakolbinger. To lead or not to lead: Endogenous sequencing in public goods games. *Economics Letters*, 108(1):93–95, 2010.

Björn Hartig, Bernd Irlenbusch, and Felix Kolle. Conditioning on what? heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics*, 55:48–64, 2015.

Karen Evelyn Hauge. Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology*, 54:35–43, 2016.

Joseph Henrich and Michael Muthukrishna. The origins and psychology of human cooperation. *Annual Review of Psychology*, 72:207–240, 2021.

Jérôme Hergueux and Nicolas Jacquemet. Social preferences in the online laboratory: a randomized experiment. *Experimental Economics*, 18(2):251–283, 2015.

Benjamin E Hermalin. Toward an economic theory of leadership: Leading by example. *American Economic Review*, pages 1188–1206, 1998.

Benedikt Herrmann, C. Thoni, and Simon Gächter. Antisocial Punishment Across Societies. *Science*, 319(5868):1362–1367, mar 2008. ISSN 0036-8075. doi: 10.1126/science.1153808. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1153808.

F. Holzmeister and R. Kerschbamer. oTree: The Equality Equivalence Test. *Journal of Behavioral and Experimental Finance*, 22:214–222, 2019. ISSN 22146350.

doi: 10.1016/j.jbef.2019.04.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2214635018302600.

Bailey R House. How do social norms influence prosocial development? *Current Opinion in Psychology*, 20:87–91, 2018.

Maximilian Hoyer, Nadège Bault, Ben Loerakker, and Frans Van Winden. Destructive behavior in a fragile public good game. *Economics Letters*, 123:295–299, 2014.

R. Mark Isaac and James M. Walker. Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *The Quarterly Journal of Economics*, 103(1):179, feb 1988. ISSN 00335533. doi: 10.2307/1882648. URL https://academic.oup.com/qje/article-lookup/doi/10.2307/1882648.

R. Mark Isaac, James M. Walker, and Susan H. Thomas. Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, 43(2):113–149, 1984. ISSN 0048-5829. doi: 10.1007/BF00140829. URL http://link.springer.com/10.1007/BF00140829.

Andrea Isoni and Robert Sugden. Reciprocity and the paradox of trust in psychological game theory. *Journal of Economic Behavior & Organization*, 167:219–227, 2019.

Toshiji Kawagoe and Yusuke Narita. Guilt aversion revisited: An experimental test of a new model. *Journal of Economic Behavior & Organization*, 102:1–9, 2014.

Harold H. Kelley and Anthony J. Stahelski. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social*

Psychology, 16(1):66–91, 1970. ISSN 0022-3514. doi: 10.1037/h0029849. URL http://content.apa.org/journals/psp/16/1/66.

Daniel Kelly. Internalized norms and intrinsic motivation: Are normative motivations psychologically primitive. *Emotion Review*, pages 36–45, 2020.

Claudia Keser and Frans van Winden. Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics*, 102(1): 23–39, mar 2000. ISSN 0347-0520. doi: 10.1111/1467-9442.00182. URL http://doi.wiley.com/10.1111/1467-9442.00182.

Oliver Kim and Mark Walker. The free rider problem: Experimental evidence. *Public Choice*, 43(1):3–24, 1984. ISSN 0048-5829. doi: 10.1007/BF00137902. URL http://link.springer.com/10.1007/BF00137902.

Martin G Kocher, Peter Martinsson, Dominik Matzat, and Conny Wollbrant. The role of beliefs, trust, and risk in contributions to a public good. *Journal of Economic Psychology*, 51:236–244, 2015.

Maria Kozlovskaya and Antonio Nicolò. Public good provision mechanisms and reciprocity. *Journal of Economic Behavior & Organization*, 167:235–244, 2019.

Roderick M. Kramer, Charles G. McClintock, and David M. Messick. Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality*, 54(3):576–582, sep 1986. ISSN 0022-3506. doi: 10.1111/j.1467-6494.1986.tb00413.x. URL http://search.ebscohost.com/login.aspx?direct=true{&}db=psyh{&}AN=1987-28122-001{&}site=ehost-live{&}scope=sitehttp://doi.wiley.com/10.1111/j.1467-6494.1986.tb00413.x.

Robert Kurzban and Daniel Houser. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102(5):1803–1807, feb 2005. ISSN 0027-8424. doi: 10.1073/pnas.0408759102. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0408759102.

Robert Kurzban, Kevin McCabe, Vernon L Smith, and Bart J Wilson. Incremental commitment and reciprocity in a real-time public goods game. *Personality and Social Psychology Bulletin*, 27(12):1662–1673, 2001.

Maria Knight Lapinski and Rajiv N Rimal. An explication of social norms. *Communication theory*, 15(2):127–147, 2005.

John O. Ledyard. Public Goods: A Survey of Experimental Research. *The Handbook of Experimental Economics*, (February):0–3, 1994. URL http://resolver.caltech.edu/CaltechAUTHORS:20171107-163442067.

M. Vittoria Levati, Matthias Sutter, and Eline Van der Heijden. Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution*, 51(5):793–818, 2007.

David K. Levine. Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, 1(3):593–622, 1998. ISSN 10942025. doi: 10.1006/redy.1998.0023.

Wim B. G. Liebrand. The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, 14(3):239–264, jul 1984. ISSN 00462772. doi: 10.1002/ejsp.2420140302. URL http://doi.wiley.com/10.1002/ejsp.2420140302.

Wim B. G. Liebrand and Charles G. McClintock. The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality*, 2 (3):217–230, sep 1988. ISSN 08902070. doi: 10.1002/per.2410020304. URL http://doi.wiley.com/10.1002/per.2410020304.

Raúl López-Pérez. Guilt and shame: An axiomatic analysis. *Theory and Decision*, 69(4):569–586, 2010.

George J Mailath and Larry Samuelson. *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.

Charles F Manski and Claudia Neri. First-and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior*, 81:232–254, 2013.

R. Mark Isaac, Kenneth F. McCue, and Charles R. Plott. Public goods provision in an experimental environment. *Journal of Public Economics*, 26(1):51–74, feb 1985. ISSN 00472727. doi: 10.1016/0047-2727(85)90038-6. URL https://linkinghub.elsevier.com/retrieve/pii/0047272785900386.

Gerald Marwell and Ruth E. Ames. Experiments on the Provision of Public Goods. II. Provision Points, Stakes, Experience, and the Free-Rider Problem. *American Journal of Sociology*, 85(4):926–937, jan 1980. ISSN 0002-9602. doi: 10.1086/227094. URL https://www.journals.uchicago.edu/doi/10.1086/227094.

David Masclet, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. Monetary and Nonmonetary Punishment in the Voluntary

Contributions Mechanism. *American Economic Review*, 93(1):366–380, feb 2003. ISSN 0002-8282. doi: 10.1257/000282803321455359. URL http://pubs.aeaweb.org/doi/10.1257/000282803321455359http://link.springer.com/10.1007/s00191-004-0212-1.

Bryan C. McCannon. Leadership and motivation for public goods contributions. *Scottish Journal of Political Economy*, 65(1):68–96, 2018. ISSN 14679485. doi: 10.1111/sjpe.12151.

Charles G. McClintock and Scott T. Allison. Social Value Orientation and Helping Behavior1. *Journal of Applied Social Psychology*, 19(4):353–362, mar 1989. ISSN 0021-9029. doi: 10.1111/j.1559-1816.1989.tb00060.x. URL http://doi.wiley.com/10.1111/j.1559-1816.1989.tb00060.x.

Charles G. McClintock and Wim B. G. Liebrand. Role of interdependence structure, individual value orientation, and another's strategy in social decision making: A transformational analysis. *Journal of Personality and Social Psychology*, 55 (3):396–409, 1988. ISSN 1939-1315. doi: 10.1037/0022-3514.55.3.396. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.55.3.396.

Charles G. McClintock and Eddy van Avermaet. Social values and rules of fairness: A theoretical perspective. 1982.

David M. Messick and Charles G. McClintock. Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, 4(1):1–25, jan 1968. ISSN 00221031. doi: 10.1016/0022-1031(68)90046-2. URL https://linkinghub.elsevier.com/retrieve/pii/0022103168900462.

J Michael. Public Goods Provision in an Experimental Environment. (June), 1985. URL http://resolver.caltech.edu/CaltechAUTHORS:20171003-133551530.

Wladislaw Mill and Maik MP Theelen. Social value orientation and group size uncertainty in public good dilemmas. *Journal of Behavioral and Experimental Economics*, 81:19–38, 2019.

Erling Moxnes and Eline Van der Heijden. The effect of leadership in a public bad experiment. *Journal of Conflict Resolution*, 47(6):773–795, 2003.

Ryan O Murphy and Kurt A Ackermann. Explaining behavior in public goods games: How preferences and beliefs affect contribution levels. *Available at SSRN 2244895*, 2013.

Ryan O. Murphy and Kurt A. Ackermann. Social Value Orientation. *Personality and Social Psychology Review*, 18(1):13–41, feb 2014. ISSN 1088-8683. doi: 10.1177/1088868313501745. URL http://journals.sagepub.com/doi/10.1177/1088868313501745.

Ryan O. Murphy, Kurt A. Ackermann, and Michel Handgraaf. Measuring Social Value Orientation. *SSRN Electronic Journal*, 6(8):771–781, 2011. ISSN 1556-5068. doi: 10.2139/ssrn.1804189. URL http://www.ssrn.com/abstract=1804189.

Ahmad K. Naimzada and Fabio Tramontana. Interdependent preferences. *Lecture Notes in Economics and Mathematical Systems*, 613:127–142, 2009. ISSN 00758442. doi: 10.1007/978-3-540-68409-1-6.

Tibor Neugebauer, Javier Perote, Ulrich Schmidt, and Malte Loos. Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology*, 30(1):52–60, 2009.

Yaw Nyarko and Andrew Schotter. An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3):971–1005, 2002.

Theo Offerman, Joep Sonnemans, and Arthur Schram. Value Orientations, Expectations and Voluntary Contributions in Public Goods. *The Economic Journal*, 106(437):817–845, jul 1996. ISSN 00130133. doi: 10.2307/2235360. URL https://academic.oup.com/ej/article/106/437/817-845/5159142.

A Yeşim Orhun. Perceived motives and reciprocity. *Games and economic behavior*, 109:436–451, 2018.

Elinor Ostrom, James Walker, and Roy Gardner. Covenants with and without a Sword: Self-Governance Is Possible. *American Political Science Review*, 86 (02):404–417, jun 1992. ISSN 0003-0554. doi: 10.2307/1964229. URL http://www.journals.cambridge.org/abstract{_}S0003055400089048.

Thomas R. Palfrey and Jeffrey E. Prisbrey. Anomalous Behavior in Public Goods Experiments : How Much and Why ? *The American Economic Review*, 87(5): 829–846, 1997. URL http://www.jstor.org/stable/2951327.

Thomas R Palfrey and Howard Rosenthal. Participation and the provision of discrete public goods: a strategic analysis. *Journal of public Economics*, 24(2): 171–193, 1984.

Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. The persuasive power of data visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2211–2220, 2014.

Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 15–03, 2015.

Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3659–3669. ACM, 2016.

David Partington and Hilary Harris. Team role balance and team performance: an empirical study. *Journal of Management Development*, 1999.

Amrish Patel and Alec Smith. Guilt and participation. *Journal of Economic Behavior & Organization*, 167:279–295, 2019.

Marco Perugini, Jonathan HW Tan, and Daniel John Zizzo. Which is the more predictable gender? public good contribution and personality. *Public Good Contribution and Personality*, 2005.

Jan Luca Pletzer, Daniel Balliet, Jeff Joireman, D. Michael Kuhlman, Sven C. Voelpel, and Paul A. M. Van Lange. Social Value Orientation, Expectations, and Cooperation in Social Dilemmas: A Meta-analysis. *European Journal of Personality*, 32(1):62–83, jan 2018. ISSN 08902070. doi: 10.1002/per.2139. URL http://doi.wiley.com/10.1002/per.2139.

Jan Potters, Martin Sefton, and Lise Vesterlund. Leading-by-example and signaling in voluntary contribution games: an experimental study. *Economic Theory*, 33 (1):169–182, 2007.

Matthew Rabin. Incorporating Fairness into Game Theory and Economics. *Amer-*

ican Economic Review, 83(5):1281–1302, sep 1993. URL https://www.jstor.org/stable/2117561{%}0A.

Katharina Rall, Margaret L. Satterthwaite, Anshul Vikram Pandey, John Emerson, Jeremy Boy, Oded Nov, and Enrico Bertini. Data visualization for human rights advocacy. *Journal of Human Rights Practice*, 8(2):171–197, 2016. doi: 10.1093/jhuman/huw011.

David G. Rand and Martin A. Nowak. Human cooperation. *Trends in Cognitive Sciences*, 17(8):413–425, aug 2013. ISSN 13646613. doi: 10.1016/j.tics.2013.06.003. URL http://dx.doi.org/10.1016/j.tics.2013.06.003https://linkinghub.elsevier.com/retrieve/pii/S1364661313001216.

Frank Renkewitz, Heather M Fuchs, and Susann Fiedler. Is there evidence of publication biases in jdm research? *Judgment and Decision making*, 6(8):870, 2011.

Ernesto Reuben, Paola Sapienza, and Luigi Zingales. Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91, 2009.

Pedro Rey-Biel. Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior*, 65(2):572–585, 2009.

M Fernanda Rivas and Matthias Sutter. The benefits of voluntary leadership in experimental public goods games. *Economics Letters*, 112(2):176–178, 2011.

Stephanie Rosenkranz and Utz Weitzel. Network structure and strategic investments: An experimental analysis. *Games and Economic Behavior*, 75(2):898–920, 2012. ISSN 08998256. doi: 10.1016/j.geb.2012.02.003. URL http://dx.doi.org/10.1016/j.geb.2012.02.003.

Paul R Sackett and Philip T Walmsley. Which personality attributes are most important in the workplace? *Perspectives on Psychological Science*, 9(5):538–551, 2014.

Selhan Garip Sahin, Catherine Eckel, and Mana Komai. An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, 1(1):100–113, 2015.

Anya Savikhin Samek and Roman M. Sheremeta. Recognizing contributors: an experiment on public goods. *Experimental Economics*, 17:673–690, 2014.

Karl H. Schlag, James Tremewan, and Joël J. Van der Weele. A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics*, 18: 457–490, 2015.

Andrew Schotter and Isabel Trevino. Belief elicitation in the laboratory. *Annual Review of Economics*, 6:103–128, 2014.

Brent Simpson. Social Values, Subjective Transformations, and Cooperation in Social Dilemmas. *Social Psychology Quarterly*, 67(4):385–395, dec 2004. ISSN 0190-2725. doi: 10.1177/019027250406700404. URL http://eprints.uwe.ac.uk/9772/http://journals.sagepub.com/doi/10.1177/019027250406700404.

D Smeesters, Luk Warlop, and Eddy Van Avermaet. Exploring the role of consistency of social value orientations : temporal stability, reciprocal cooperation, and forgiveness. *DTEW Research Report 0238*, pages 1–34, 2002. URL https://lirias.kuleuven.be/bitstream/123456789/224412/1/OR{_}0238.pdf.

D Smeesters, Luk Warlop, Eddy Van Avermaet, Olivier Corneille, and Vincent Yzerbyt. Do not prime hawks with doves: The interplay of construct activation and consistency of social value orientation on cooperative behavior. *Journal of Personality and Social Psychology*, 84(5):972–987, 2003. ISSN 1939-1315. doi: 10.1037/0022-3514.84.5.972. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.84.5.972.

Alexander Smith. Estimating the causal effect of beliefs on contributions in repeated public good games. *Experimental Economics*, 16(3):414–425, 2013.

Joel Sobel. Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, 43(2):392–436, may 2005. ISSN 0022-0515. doi: 10.1257/0022051054661530. URL http://qjmam.oxfordjournals.org/content/41/4/503.abstracthttp://pubs.aeaweb.org/doi/10.1257/0022051054661530.

Joep Sonnemans, Frans Van Dijk, and Frans Van Winden. On the dynamics of social ties structures in groups. *Journal of Economic Psychology*, 27:187–204, 2006.

Christian Thöni and Stefan Volk. Conditional cooperation: Review and refinement. *Economics Letters*, 171:37–40, 2018.

Eric van Dijk, David De Cremer, and Michel J.J. Handgraaf. Social value orientations and the strategic use of fairness in ultimatum bargaining. *Journal of Experimental Social Psychology*, 40(6):697–707, nov 2004. ISSN 00221031. doi: 10.1016/j.jesp.2004.03.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0022103104000289.

Frans Van Dijk, Joep Sonnemans, and Frans Van Winden. Social ties in a public good experiment. *Journal of Public Economics*, 85:275–299, 2002.

Paul A. M. Van Lange. Confidence in expectations: A test of the triangle hypothesis. *European Journal of Personality*, 6(5):371–379, dec 1992. ISSN 08902070. doi: 10.1002/per.2410060505. URL http://doi.wiley.com/10.1002/per.2410060505.

Paul A. M. Van Lange. the Pursuit of Joint Outomes and Equality of Outcomes. *Journal of Personality and Social Psychology*, 77(2):337–349, 1999.

Paul A. M. Van Lange and Wim B. G. Liebrand. On perceiving morality and potency: Social values and the effects of person perception in a give-some dilemma. *European Journal of Personality*, 3(3):209–225, sep 1989. ISSN 08902070. doi: 10.1002/per.2410030306. URL http://www.interscience.wiley.com/jpages/0890-2070/http://www.wiley.comhttp://doi.wiley.com/10.1002/per.2410030306.

Paul A. M. Van Lange and Wim B. G. Liebrand. Social value orientation and intelligence: A test of the goal prescribes rationality principle. *European Journal of Social Psychology*, 21(4):273–292, jul 1991. ISSN 00462772. doi: 10.1002/ejsp.2420210402. URL http://doi.wiley.com/10.1002/ejsp.2420210402.

Paul A. M. Van Lange, Christopher R. Agnew, Fieke Harinck, and Gemma E. M. Steemers. From game theory to real life: How social value orientation affects willingness to sacrifice in ongoing close relationships. *Journal of Personality and Social Psychology*, 73(6):1330–1344, 1997a. ISSN 00223514. doi: 10.1037/0022-3514.73.6.1330.

Paul A. M. Van Lange, Ellen M. N. De Bruin, Wilma Otten, and Jeffrey A. Joireman. Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73

(4):733–746, 1997b. ISSN 1939-1315. doi: 10.1037/0022-3514.73.4.733. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.73.4.733.

Paul A. M. Van Lange, René Bekkers, Theo N. M. Schuyt, and Mark Van Vugt. From Games to Giving: Social Value Orientation Predicts Donations to Noble Causes. *Basic and Applied Social Psychology*, 29(4):375–384, nov 2007. ISSN 0197-3533. doi: 10.1080/01973530701665223. URL https://www.tandfonline.com/doi/full/10.1080/01973530701665223.

Paul A. M. Van Lange, Jeff Joireman, Craig D. Parks, and Eric Van Dijk. The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, 120(2):125–141, mar 2013. ISSN 07495978. doi: 10.1016/j.obhdp.2012.11.003. URL http://dx.doi.org/10.1016/j.obhdp.2012.11.003https://linkinghub.elsevier.com/retrieve/pii/S0749597812001276.

Christoph Vanberg. Why do people keep their promises? an experimental test of two explanations 1. *Econometrica*, 76(6):1467–1480, 2008.

Daniel Villatoro, Giulia Andrighetto, Rosaria Conte, and Jordi Sabater-Mir. Self-policing through norm internalization: a cognitive solution to the tragedy of the digital commons in social networks. 2015.

Marie Claire Villeval. Public goods, norms and cooperation. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing, 2020.

Nathaniel T Wilcox, Nick Feltovich, et al. Thinking like a game theorist: Comment. *University of Houston Department of Economics working paper*, 2000.

Toshio Yamagishi. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1):110–116, 1986. ISSN 1939-1315. doi:

10.1037/0022-3514.51.1.110. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.51.1.110.