Essays on Social Norms and Cooperation

Jiawen Li

Doctor of Philosophy

University of York

Economics

SEPTEMBER 2014

Abstract

This thesis is a collection of experimental and theoretical studies on social norms and cooperation.

The first two chapters focus on effects of social norms on people's behaviour in strategic situations. The Golden Rule is first studied in an ultimatum bargaining experiment. The results show that while most people follow the Golden Rule in the ultimatum game situation, experience and feedback of playing the opposite role has an important effect on golden-rule behaviour. Then the link between people's expectation of social norms and their own behaviour is studied in an experiment of a trust game. Only about half of the subjects show a consistent behaviour according to their own expected norm. Moreover, experience and feedback has asymmetric effects on the behaviour of trustors and trustees.

In next two chapters, the way people cooperate and how to sustain a more efficient cooperative result are studied by using both theoretical and experimental methods. I first experimentally explore the mechanisms that make people more willing to cooperate and increase the overall welfare in a public goods game. Then I theoretically study a well-established cooperative solution for the bankruptcy problem and design a non-cooperative game that gives the solution as the unique sub-game perfect equilibrium outcome.

Contents

List of Figures

List of Tables

Acknowledgements

This work is made possible by all the amazing people who give me help and support generously.

I am eternally in debt to my supervisor Yuan Ju, who is the best supervisor anyone can ever ask for. He is the one who opened the door to an academic career for me more than four years ago and since then constantly keeps me on the right path. With sparks of genius, he shows me the way of being a theorist and nurtures my interest in experimental economics. His attitude towards work and life is the priceless lesson for me.

My thesis advisory group member Professor John D. Hey and Dr. Anindya Bhattacharya have been tremendously helpful during my study. Not only did they provide valuable comments and suggestions on my work, but also the regular meetings with them kept my progress in check.

I am truly grateful for having Miguel Costa-Gomes as my coauthor and good friend. His work ethics and constant pursuing of perfection has a profound influence on my research. A very special thanks goes to the people of GLOPEII at Waseda University, Tokyo, Japan, especially Yukihiko Funaki and Róbert F. Veszteg, who are a lot of fun to work with.

I had the great pleasure to meet many researchers, who are kind and patient and gave me helpful advices. I also had the fortune to make many close friends through the journey. They are my constant inspiration. To all of them I owe my greatest appreciation.

I am grateful for all my friends, Xueqi Dong, Yu Wang, Yuanyuan Xie and Chi Zhang, who are always there for me. I also want to thank Jens Gudmundsson, who, by comparison, constantly makes me look lazy and stupid.

Last but not least, I am grateful for my parents who support my choices unconditionally. To them, I owe my everything.

Declaration of Authorship

I, Jiawen Li, declare that this thesis titled, 'Essays on Social Norms and Cooperation' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made reference to that.

To my parents

“All the world is made of faith, and trust, and pixie dust.”

J.M. BARRIE, Peter Pan

Chapter 1

Introduction

1.1 Background

Game theory is "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers", as stated on the first page of Myerson's canonical game theory textbook (Myerson, 1991). The analysis of games, that is the strategic interactive situations, shows what completely rational people would do in specific situations.

Real life strategic interactions are ubiquitous. In the Istanbul grand bazaar, a customer and a shopkeeper bargain over the price of a rug. A venture capital fund decides on how much to invest on a startup company. In a community, residents decide how much to donate to a project of building a park. When a firm goes bankrupt, creditors negotiate to decide how the liquidation should be divided. Game theoretical analysis of above situations would tell us if people were all trying to serve their own best, what they would do. A closer look at the above examples would naturally lead to the question of what people actually do. Unsurprisingly, people in real life do not always behave consistently with the prediction of game theory, not only because people are not completely rational, as assumed in the game theoretical analysis, but also because being social creatures, our decisions are influenced by considerations other than best-serving our own interests.

If relaxing the assumption of rationality, we would enter the world of behavioural game theory. As a branch of behavioural economics, it "uses psychological regularity to suggest ways to weaken rationality assumption and extend theory" (Camerer, 2003). By utilising controlled experimental environment, behavioural game theory answers the question of how people behave under certain circumstances. Moreover, it suggests possible other factors beyond self-utility maximising that might also influence people's decision making.

This thesis is a collection of essays on different aspects of game theory utilising both theoretical and experimental approaches. The two approaches complement each other in order to provide a comprehensive insight on the topics.

Two general aspects that I will explore are the influences of social norms and cooperation in strategic situations. Although distinctive in forms, social norm and cooperation are often inseparable. As the unresolved issues in experimental game theory, different forms of social norms have been studied in order to explain the cooperative behaviour in games. Theoretical models have been proposed to incorporate social norm in agents decision making in strategic situations. Cooperation in modern society is mainly supported by social norms. Theoretical analysis of cooperation still assumes the rationality of players and more often than not, enforceable agreements among players (as in *cooperative game theory*), while the experimental study of cooperation focuses on the rationale behind the cooperation and explores ways to reinforce social norms and improve the cooperation in strategic situations.

In this thesis, I first focus on two important social norms and phenomena, the Golden Rule and Trust related issues. Then, I experimentally explore possible ways to improve the cooperation level in a specific setting and theoretically study how a mechanism could be designed to achieve a cooperative solution without a binding agreement.

1.2 Social Norms

In many interactive situations with two or more agents, there is a conflict of interests. Whenever there is such conflict, people involved need to negotiate to achieve an agreement, which in most cases means to sacrifice part of their own welfare in order to achieve a joint gain. This is when social norms kick in. People in strategic situations or when there is a conflict of interests, care about not only their own payoffs, but also many other things like other people's well-being, how other people would look at them if people can observe their behaviour, or whether they conform to their own perception of the social norms even when other people cannot observe their behaviour. We may also apply social norms unconsciously, but still be able to explain our behaviour in terms of beliefs (Bicchieri, 2006). The norms may have affected how we think we should behave under certain circumstances, or how we expect others to behave in that situation, or even how we should behave because of others' expectation on us.

Experimental studies observe how these social norms influence people's decisions in different interactive settings. They help us to better understand the reasoning process in specific games, as well as provide inspirations for the development of theories.

Here I focus on two specific social norms: the Golden Rule and the trust.

1.2.1 Golden Rule

As one of the most universally studied maxims, the "Golden Rule" appears in almost every ethical tradition in different forms (Blackburn, 2001). The concept of "treat others in a way that you yourself would like to be treated" appears long before the term of "Golden Rule". As early as in the *Analects of Confucius*, there is teaching of *"Never impose on others what you would not choose for yourself"*. Similar teachings also appear in other major religions, including Hinduism, Buddism, Taoism and Zoroastrianism (Spooner, 1914). It is so deeply rooted in people's everyday life, that consciously or unconsciously in decision making, the golden rule sneaks into our mind and ask "Would I want the other people to do the same to me?". Bergstrom (2009) describes four categories of the golden rule: "love thy neighbour, do-unto-others, negative do-unto-others, Kantian Categorical Imperative". In Chapter 2, I will explore the implication of "negative do-unto-others" in a specific strategic setting. In order to observe the Golden Rule in a situation of conflicts, the ultimatum game[1] is chosen as a benchmark game. The simple nature of the game makes it easy for the subjects to understand the game situation, while preserving the property of conflicting interests between two players. More specifically, one of the players, the proposer, makes an offer of how to divide a certain amount of money; the other player, the responder, either accepts or rejects the offer. In the case of acceptance, they each get the amount as proposed. In the case of rejection, they both get nothing.

The main question that I will try to answer is "Do people's behaviour conforms with the teaching of the Golden Rule?". Although the Golden Rule has been studied extensively in many fields, this is the first study from an experimental economics point of view. The study quantitatively defines the proportion of people who conform to the Golden Rule under certain circumstances. It also sheds light on how experience and feedback from "walking in the opponents' shoes" could shape people's behaviour.

1.2.2 Trust and Expected-Norm

Trust is often seen as an important form of social capital, which reduces the transaction cost in an economy and increases the economy development (Zak and Knack, 2001). In the context of strategic situations, however, an act of trust would place the trustor in a vulnerable position where he willingly relinquishes the control of his own payoff which would then depend on what the trustee does.

The trust game, initially designed by Berg et al. (1995), provides an efficient way to study trust and trustworthiness. It is a two-person game, involving a trustor and a trustee. The trustor decides how much out of a

[1] The experimental study of the ultimatum game popularised since the seminal work by Güth et al. (1984). See Güth and Kocher (2014) for a recent survey.

certain amount of money to send to the trustee. The amount sent will be tripled on the way to the trustee. The trustee after receiving the money decides how much to return to the trustor. The amount sent is designed to be a measurement of the level of trust, while the amount returned measures the level of trustworthiness.

I look at trust (and trustworthiness) from a different angle in Chapter 3. Different from other studies the trust game, I link a subject's expectation of his opponent to his own behaviour in the opponent's position. By looking at the effect of the expectation on people's decision making and relating the expectation to their own behaviour in the same situation, I answer the basic question of how people's behaviour fits into their own perception of the norm that the others would follow.

1.3 Cooperation

As discussed before, completely rational agents may not like cooperation when there is a conflict of interests, as it requires a sacrifice of their own interests. Nevertheless, in reality, people do sometimes cooperate in order to increase the overall benefit. In chapter 4 and 5, I explore possible ways to improve the cooperative behaviour, experimentally in a public-goods provision situation and theoretically for bankruptcy problems.

1.3.1 Public Goods Game

The general concept of a public-goods situation is when people can contribute part of their own endowments to a public project which would in return benefit themselves. The more the other people contribute the more benefit one will get. The aggregate contribution represents the overall welfare of the whole society. The problem here is to fight the incentive of free-riding. People need to sacrifice their own payoffs in order to generate a greater welfare overall. The equilibrium outcome is that no player contributes and no public good is provided. It has long been shown that people are willing to contribute in the game, but not all of their endowment. Typically, they contribute about 50% of their endowments. (See Ledyard, 1995 and Camerer, 2003 for surveys.)

Effort has been made in trying to find mechanisms to increase the contribution and reduce the free-riding. Popular mechanisms include repetition, punishment, pre-game communication and etc. One mechanism that recently caught much attention is the institution formation. People are given the choice of joining an institution which would force its members to contribute at a certain level. The mechanism readily gains popularity due to the resemblance of the negotiations in EU and on international environmental agreements. The study in Chapter 4 builds on the institution formation mechanism by Kosfeld et al. (2009), and investigates the effects of different voting systems within the institution. A modified mechanism to improve the efficiency in the public goods game is proposed and compared to other possible mechanisms.

1.3.2 Bankruptcy Problems

Theoretical study of the cooperative games is typically divided into axiomatic and strategic approach. While the axiomatic approach characterises the solutions of cooperative game with sets of axioms, the strategic method approaches a cooperative game solution by designing a non-cooperative game whose equilibrium outcome is the desired solution. Such method has been used to build the non-cooperative foundation for cooperative solutions, and helps us to have a better understanding of the cooperative solutions. For example, the non-cooperative games to the core (see, for example, Perry and Reny, 1994 and Serreno and Vohra, 1997) and Shapley value (see, for example, Pérez-Castrillo and Wettstein, 2001) provide intuitive game procedures for those solutions to be achieved as non-cooperative equilibria.

I focus on a fair division problem called the bankruptcy problem. A bankruptcy problem describes a general situation when several creditors who each entails to a certain part of an estate decide on how the estate should be divided. The sum of all the entitlement is larger than the whole estate. There are several classical solutions to this problems, for example, the proportional rule, the constrained equal awards rule and the constrained equal losses rule. Talmud rule, originated from the Talmud, was first formulated by Aumman and Maschler. It possesses several desirable properties of the fair division problem.

The bankruptcy problem has been studied as a cooperative game, where the Talmud rule is found to coincide with the Nucleolus of the corresponding bankruptcy game. Chapter 5 looks at the bankruptcy problem from a non-cooperative point of view. A strategic game is designed such that its unique sub-game perfect equilibrium outcome coincides with the Talmud solution of the bankruptcy problem. There are several games proposed with the same goal in mind, for example, by Serrano (1995) and Dagan et al. (1997). The game proposed here, however, does not rely on the consistency property of the solution which is different from the existing studies. It also cherishes the cake-cutting mechanism in fair division.

Chapter 2

An Experimental Study of the Golden Rule[1]

Never impose on others what you would not choose for yourself.

– CONFUCIUS

2.1 Introduction

This chapter addresses a general, yet, open issue common to most societies, cultures, and religions: whether agents' behaviour in some economic environments conforms to the Golden Rule. The Golden Rule is a central concept about human interaction that has a long history, and, as put by Blackburn (2001), "can be found in some form in almost every ethical tradition".

The form of the Golden Rule most frequently referred to by a wide range of cultures and religions states that one should not treat others in ways that are not agreeable to oneself: in The Analects of Confucius, it is said "Never impose on others what you would not choose for yourself"; in Buddhism, it is said "Hurt not others in ways that you yourself would

[1]This chapter is based on *Role-Reversal Consistency– An experimental Study of the Golden Rule*, a joint work with Miguel Costa-Gomes and Yuan Ju.

find hurtful"; according to the Talmud, "That which is hateful to you, do not do to your fellow"; according to the Zoroastrian Shayast-na-Shayast, "Whatever is disagreeable to yourself do not do unto others"; in Islam, the Golden Rule is implicitly expressed in the Qur'an, but explicitly stated in the Hadith.[2] We confine our study to this form of the Golden Rule (and henceforth, we have it in mind when we refer to the Golden Rule).

The Golden Rule is an ethical norm or a moral principle that approaches strategic situations in a distinctive way: first, it tells the agent to think about the opposite role; second, suggests to the agent to think how she would behave in the opposite role rather than how an abstract opponent would behave in that role (in other words it replaces the agent's belief about the behaviour of the agent who is her opponent as is standard in game theoretic reasoning, with what her own behaviour in the role of her opponent would be; however, note that under projection bias (Allport, 1924; Krueger and Acevedo, 2005) the former and the latter coincide, as the agent's belief about her opponent's behaviour mimics the agent's own behaviour in that role); finally, it tells the agent to take her hypothetical behaviour in the opposite role into account in a particular way when deciding what to play in her current role.

The idea that an agent should consider how she should play the different roles in a game, that is key to the Golden Rule, has found its way into the economics literature. Gerchak and Fuller (1992) offer a first theoretical analysis of the dissolution of business partnerships that include buy-sell clauses in contracts. According to such clauses, one owner proposes a buy-sell price and the other owner is compelled to either purchase the proposer's shares or sell her own shares to him at the proposed price. This mechanism embodies the spirit of the Golden Rule, as the price at which one party proposes to buy out the other is also the price at which he would have to sell his stake to her. A setting with symmetric agents illustrates this

[2] The Golden Rule has other interpretations. Its positive form states that one should treat others as one would like others to treat oneself, as appears in the Bible "So whatever you wish that others would do to you, do also to them, for this is the Law and the Prophets." Bergstrom (2009) describes four versions of the Golden Rule: "Love-thy-neighbor", "Do-unto-others", "Negative do-unto-others" (the negative form we allude to), and Immanuel Kant's Categorical Imperative.

in a vivid way: the optimal proposed price by any of the two agents can only lie in between the agent's minimum willingness to sell and maximum willingness to buy, as otherwise she will make a loss because the other agent will sell his stake to her, if the price is higher than the latter, or will buy her share, if the price is lower than the former. Any offer that the agent would find hurtful to herself if she were playing the other role, should not be made, as it would be turned down by the other party, and would then hurt her. Thus, a mutually beneficial deal (also Pareto efficient) is achieved only when both agents act according to *treat no one in ways that they find not agreeable to themselves*.

Another mechanism whose game theoretic solution (in this case its unique subgame perfect equilibrium) agrees with the Golden rule is Brams and Taylor's (1996) divide-and-choose procedure in fair division. Very recently, Alger and Weibull (2013) study the role that the Golden Rule plays in the evolutionary stability of preferences in the context of assortative matching.

The main focus of this chapter is to experimentally test the behavioural implications of the Golden Rule, i.e., of *treat no one in ways that are not agreeable to yourself*. We do this by asking subjects to play both roles of a (modified) ultimatum bargaining game, in which each subject in each role faces multiple independent opponents simultaneously.

In an ultimatum game, the proposer suggests a division of an amount of money which the responder either accepts or rejects. If the responder accepts, the money is divided between both players as proposed, otherwise both get nothing. In this game, the Golden Rule has an intuitive interpretation: an agent playing the responder's role accepts the offer she makes as a proposer. To be more precise, if an agent, when playing the role of a responder (i.e., when dealing with how others treat her), accepts the offer she would make as a proposer (i.e., when treating others), then that implies she does NOT treat others in ways she finds not agreeable to herself, thereby conforming to the Golden Rule. We call such agent *role-reversal consistent*. On the contrary, if an agent, when playing the responder's role, rejects the offer she would make as a proposer, then that implies she DOES treat others in ways she finds not agreeable, therefore violating the Golden Rule.

We use the term role-reversal consistent, because in our study each subject plays different subjects when playing the game's two different roles (which is role-reversal), which is different from each subject playing the same subject twice, once in one role, the other time in the other role (which is role-switching). We believe that the principle of *Never impose on others what you would not choose for yourself* reflects better the way people should behave when interacting with different people, than when interacting with the same person repeatedly. Furthermore, role-reversal eliminates reciprocity across games (apart from anonymous reciprocity) likely to arise in a role-switching situation.

The goal of our study is neither to dispute the well known stylized facts of ultimatum game experiments (for a survey see Camerer (2003), among others) nor to question any of the models of other regarding preferences that center their attention on distributional preferences and/or reciprocity that purport to explain behaviour in the ultimatum and many other games. Instead, we simply study whether subjects' behaviour in this game is role-reversal consistent and identify variables that have an effect on such consistency.

Subjects first play one of the roles of the ultimatum game, being unaware they will play the opposite role later. In one treatment we use the direct-response method (henceforth, shortened to response method or response treatment), and in the other treatment we use the strategy vector method (henceforth, shortened to strategy method or strategy treatment).[3] Next, subjects play the opposite role under the same method as before, being aware that they are playing people they have not played before. In the response treatment, subjects get feedback during and in between the two

[3]While the term strategy (vector) method is uniformly used to describe a protocol where a player has to specify an action for each of her information sets, the protocol where players are only asked to choose an action for information sets on the path of play is referred to in different ways, such as "direct-response", "sequential decision", etc. Others (e.g., Brandts and Charness (2000)) refer to the two protocols as "cold" and "hot". For a survey of similarities and differences in players' behaviour under the two methods, see the recent survey by Brandts and Charness (2011).

games, while in the strategy method treatment they do not, so that the subject's elicited complete strategy profile (i.e., her choices under both roles) cannot possibly be influenced by the feedback she could receive in the role she plays first. We constrain subjects' offers to whole amounts in sterling (GBP) so that: i) we can use the strategy method for the responder's role in one treatment; ii) to increase the probability that a subject when playing the responder role is made the offer she herself makes as a proposer in the response treatment; iii) to elicit a proposer's beliefs about the probability of acceptance of each possible offer.

Our main findings are as follows.

First, we find that overall 82.01% of the subjects are role-reversal consistent. However, the response method, where subjects receive feedback about the reactions of others to the way they treat them or how others treat them, yields a higher level of role-reversal consistency, 92.97%, than the strategy method, 72.67%, where such feedback is suppressed.

Second, regression analysis suggests that the high rate of role-reversal consistency is not the result of strategic play in the role of the proposer, in the sense of the subject choosing the offer that maximizes her expected monetary earnings, given her beliefs about the responder's behaviour.

Third, role-reversal consistent subjects are more accurate predictors of the actual probability with which each offer is accepted than inconsistent subjects. Role-reversal consistent subjects also earn more money in the experiment than inconsistent ones.[4]

Fourth, we do not find that any of the demographic, socio-economic or cultural variables we collect data on has a significant effect on the level of role-reversal consistency, although the sample is large (each role is played by 300 subjects) for a laboratory experiment.

Finally, we add to the literature that compares behaviour under the strategy and the response methods. As we explain below, our design makes

[4]We also find that subjects who first play the role of proposes are more likely to be role-reversal consistent than subjects who first play the role of responders. However, this effect is only statistically significant in the treatment using the strategy method.

a methodological contribution because it sharpens the comparison of one-shot behaviour of the second mover in a two-person two-move game under the strategy and the response methods. Furthermore, our data suggests that first movers (i.e., proposers) behave differently in the two treatments not because their beliefs about second movers' (i.e., responders) behaviour are different, but because their preferences (or within the context of our study, the way they apply the Golden Rule) in the two treatments seem to be different.

2.2 Related Literature

Earlier papers have asked subjects to play both roles of a two-person game under a variety of protocols, different from ours. Güth, Schmittberger and Schwarze (1982) describe a treatment where each subject simultaneously decides how much to offer as a proposer and the minimum offer she accepts as a responder. They find that 86.49% of the subjects are, in their own terminology, "consistent", because the sum of their offer and the minimum acceptable offer is smaller than the amount to be divided.[5]

In a later study, Oxoby and McLeish (2004) ask subjects to specify a complete strategy profile in the ultimatum game, both writing down the offer they would make as a proposer and answering whether they would accept or reject each feasible offer (they use a $10 pie, and restrict offers to whole dollar amounts). They compare subjects' aggregate behaviour in this treatment with that in another treatment where subjects are assigned one of the two roles and play an ultimatum game using the direct-response method, i.e., with the proposer making an offer that is conveyed to the responder who either accepts it or not. Blanco, Engelmann and Normann (2011) also ask subjects to play both roles simultaneously in the ultimatum game using the strategy method. Neither of the studies analyzes whether a subject would accept her own offer or considers the direct response method.

[5] Although role-reversal was first used in the context of the ultimatum game, its use has spread to other games and to issues as varied as the inferring of subjects' distributional preferences, as in Charness and Rabin (2002).

More recently, Chai, Dolgosuren, Kim, Liu and Sherstyuk (2011) use role-reversal in one of their treatments. However, their study does not address the issue of subjects' role-reversal consistency. Furthermore, their design is not suited to understand the effects that feedback, or the order of play of the roles have on consistency. They run two treatments (which they call one-role and two-roles treatments) with several games, among them the ultimatum game. In their one-role treatment subjects either play the role of proposer or responder. In their two-roles treatment, they use a role-reversal protocol in which subjects play the proposer role first, and next specify a cut-off strategy for the role of responder. Their main aim is to correlate behaviour with attitudinal responses. They find that the average offer is the same in the one-role and two-roles treatments and also that the order according to which subjects play the two roles does not affect the average offer made as a proposer.

Our study differs from all the studies above. We use the response method (in addition to the strategy method), which reflects the conflicting nature of the strategic situation our study depicts more naturally, even if it makes it harder to test role-reversal consistency, as explained below. We do not force responders to report a cut-off strategy, because we want to control for the possibility that could cue subjects to choose an offer and a cut-off strategy that sum up to an amount not larger than the pie. Subjects play one role at a time, rather than both roles simultaneously, as real-life situations are better described through sequential play of opposite roles. We eliminate the 50-50 split of the pie outcome because its focalness might nudge subjects' behaviour to be role-reversal consistent. We also elicit the proposer's beliefs about the probability of acceptance of each offer, as we aim to understand whether subjects' beliefs about the responder's role reflect the way they play that role. The data we collect allows us to answer whether subjects' behaviour is role-reversal consistent and to identify factors that influence such behaviour.

At a conceptual level, the idea closest to ours is that of Blanco, Celen and Schotter (2011). They introduce and formalise the notion of "blame-freeness", which defines one's kindness. Their concept relies on role-reversal, because a subject's judgement whether her opponent is kind or unkind to her, requires her to compare the strategy played by her opponent with the strategy she would herself play in her opponent's role (if she would have acted in a more unkind manner than her opponent actually does, then her opponent is blame-free; otherwise, her opponent is blameworthy). Note that blame-freeness and role-reversal consistency are different notions. While "blame-freeness" relies on comparing an opponent's role with one's behaviour in that role, role-reversal consistency is only about comparing one's behaviour in both roles. Furthermore, we add that Blanco et al.'s work formulates a general notion of "blame-freeness" that can be applied to any game, and experimentally test it using tournament games where subjects play both advantaged and disadvantaged roles with different opponents, finding support for it. We simply apply the idea of role-reversal consistency to one game and investigate if behaviour can be interpreted according to it.

2.3 Experimental Design

Our experimental design tweaks the standard version of the ultimatum game to test whether a subject does not treat others in ways that she does not find agreeable to herself, which we call "role-reversal consistency". To conduct this test we ask subjects to play both roles of the ultimatum game. To avoid inadvertently nudging subjects to simultaneously think how they would play both roles, subjects are asked to play one role at a time, not the two roles simultaneously, unaware that they will be asked to play the other role next.

In our games players bargain over £7, but offers can only be made in whole (i.e., integer) sterling amounts (£0, £1, . . . , £7). Thus, the pie cannot be split evenly, an outcome very often observed in typical ultimatum game experiments. The availability of this outcome could yield a high level of role-reversal consistent behaviour, but would confound such explanation for behaviour with all the other explanations (ranging from pure inequality aversion to such outcome being the focal point in a bargaining situation) that have been proposed for it.[6] To avoid this confoundedness we only allow unequal splits of the pie, which sharpens the test of role-reversal consistency.

The experiment has two treatments: one uses the response method, the other uses the strategy method. In both treatments each subject is first assigned either the role of proposer or responder. In the response treatment, the proposer's offer is conveyed to the responder, who either accepts or rejects it. The proposer is informed about the responder's decision. Next, each subject plays the opposite role. In the strategy treatment, the proposer makes an offer that is not conveyed to the responder. The responder decides whether to accept or reject each of the feasible offers. Subjects are not told the outcome of the game, and play the opposite role next.

Under the strategy method, it is always possible to test whether a subject is role-reversal consistent, because in the role of the responder she is asked to accept or reject each of the feasible offer amounts, and, therefore, also the amount she offers in the proposer's role.

Under the response method, we can only test whether a subject is role-reversal consistent when in the role of a responder she is offered the amount she offers as a proposer. Since one of the stylized facts in the ultimatum game is that low and very low offers are observed infrequently, the probability that the role-reversal consistency of subjects making such offers can be tested with the ultimatum game can be quite low. Thus, in order to maximize the probability that we can test a subject's role-reversal consistency under the response method, proposers make an offer that is sent simultaneously to multiple responders, with responders receiving simultaneous offers from multiple proposers. For the sake of comparability, we also use this feature in the strategy treatment.

In our experiment subjects interacted with each other only via z-tree's computer interface (Fischbacher (2007)). Their identity and experimental ID number were kept strictly confidential from each other throughout and

[6]We thank Matthias Sutter for highlighting this possible confoundedness, which ultimately led us to drop the equal split outcome from the set of feasible offers. Güth, Huck and Muller (2001) compare behaviour between mini ultimatum games with and without the equal split outcome.

after the experiment to ensure anonymity. Each session had thirty subjects (selected from the undergraduate student population at a UK university and from many different majors with the exception of Economics to rule out subjects who have been exposed to game theory) who were randomly divided into three groups of ten, A, B and C. They were told that the session had three independent parts, and that each group would participate in two parts, interacting with a different group in each part. Subjects were not described the decision situation they would face in a part before they got to it.

In part I, group A subjects played group B subjects, the former as proposers, the latter as responders, while group C subjects were passive. In part II, group A subjects became responders and played group C subjects who were proposers, and group B subjects were passive. In part III, group B subjects became proposers and played group C subjects who became responders, and group A subjects were passive. The passivity of one group in each part reinforced to the subjects in each group the fact that they would only interact once with the subjects from each of the other two groups. This rotation of the groups suppresses reciprocity-driven behaviour across games. The structure of a session's different parts is summarized in Table 2.1 and in Figure 2.1 (A).

TABLE 2.1: The structure of the experiment

	Group	Role	Previous Role
Part I	A	Proposer	None
	B	Responder	None
Part II	C	Proposer	None
	A	Responder	Proposer
Part III	B	Proposer	Responder
	C	Responder	Proposer

In each part, each proposer made an offer that was sent to all 10 responders. Accordingly, each responder received 10 offers, one from each proposer, as illustrated in Figure 2.1 (B) and (C).

In the response treatment, each responder received the 10 offers all at once. She then had to decide whether to accept or reject each of them, and submitted all her decisions at the same time. This allowed her to both accept and reject offers of the same amount.[7]

In the strategy treatment, each responder was asked to decide whether she would accept or reject any of the feasible eight offer amounts for each of the 10 offers. She was neither constrained to choose a cut-off strategy (e.g. can reject high offers and accept low ones) nor to accept or reject a particular (hypothetical) amount for all 10 proposers. In other words, when a subject played the responder role, she could either use the same or different strategies to play the ten different proposers. Not constraining responders to use cut-off strategies provides a clean comparison of their behaviour between the two treatments.[8]

In both treatments, each proposer, after making her offer (but before the game's outcome is revealed in the response treatment), states her belief about the responder's conditional probability of acceptance for each feasible offer (i.e., for all whole amounts between £0 and £7). Subjects state their beliefs before they play the opposite role, which happens in a later part of the experiment. We use a quadratic scoring rule to elicit such beliefs, which is incentive compatible under the assumption of risk-neutral expected monetary earnings maximizing behaviour.

We determined subjects' earnings at the end of the last part of the experimental session (to suppress wealth effects as much as possible). We paid subjects for playing both roles and for the accuracy of their stated beliefs. For each subject we paid: i) the outcomes of two randomly chosen responses to her offer as a proposer (see the thick arrows in Figure 2.1 (B)); ii) the outcomes of her decisions as a responder to two randomly chosen

[7] Although there is evidence that a responder does not always accept or reject a particular amount, it comes from experiments where the ultimatum game is played repeatedly (either with same proposer or with different proposers). In such experiments this behaviour can be explained by the responder dynamically adjusting her behaviour given the history of offers she receives to either teach proposers to increase their offers or because she learns to accept lower offers. Our design rules out such explanations. In our design, the observation that a responder accepts some while rejecting other offers of the same amount is either evidence of her indifference (if that happens only for one amount), or of her choice being stochastic (if the responder acts in that way for different offer amounts).

[8] In addition, a few past studies (see Roth, Okuno-Fujiwara, Prasnikar and Zamir (1991)) document that some subjects reject very high offers, which reinforces the appropriateness of eliciting an unconstrained strategy.

offers (see the thick arrows in Figure 2.1 (C)); iii) the accuracy with which she estimates that each of two offers randomly selected from all ten offers made by the proposers in her group is accepted by one randomly chosen responder from the opposite group, with a maximum of £1 per estimate.[9]

The average payment (excluding a show-up fee of £3) was £13.7 in the response treatment and £13.0 in the other treatment. The highest and lowest earnings in the response treatment were £20.0 and £4.8, respectively, while the figures for the strategy treatment were £18.9 and £1.0, respectively.

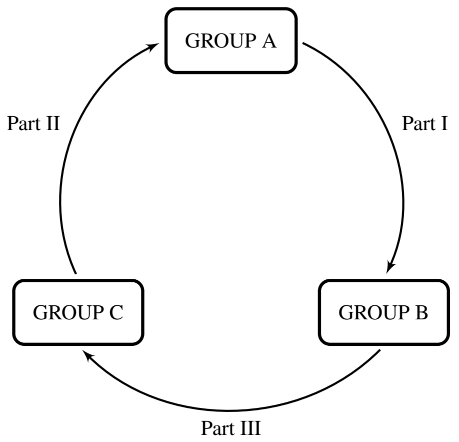
At the end of the session subjects completed a questionnaire, after which they collected their monetary earnings.[10] Each treatment had 5 sessions, and the whole experiment had 300 participants.

2.4 Data Analysis

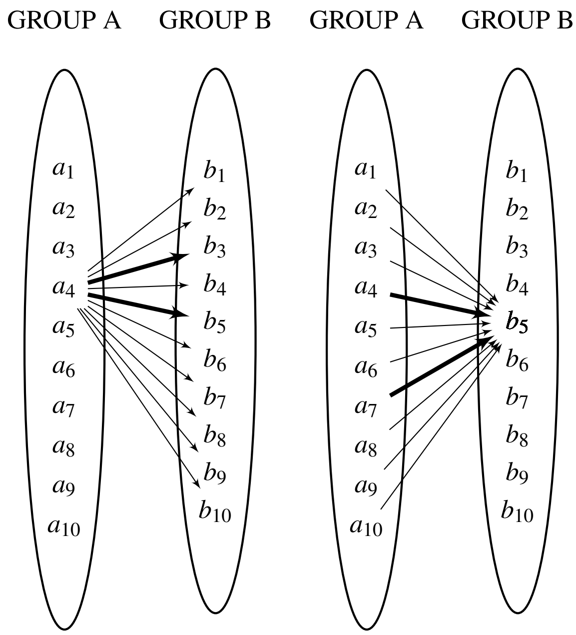
In this section, we present the data analysis and our findings. We start by providing a summary of the offers subjects made as proposers, their stated beliefs about the responder's behaviour, and their behaviour as responders. Next, we look at each subject's behaviour in both roles and check whether it is role-reversal consistent. Finally, we try to identify some factors that affect the probability that a subject is role-reversal consistent.

[9]This reward mechanism deals with hedging because it yields a very low (4%) probability that the same outcome of the subject's interaction with an opponent is chosen to reward her decision as a proposer and her stated beliefs. In addition, so far there is little empirical evidence that such concerns matter (see Blanco, Engelmann, Koch and Normann (2010)). Unlike what would happen with role-switching, in our design subjects would find it hard to maximize the minimum earnings from playing both roles, as they face different opponents in each role and are paid for just two (out of ten) randomly chosen interactions in each role at the end of the session.

[10]The questionnaire has questions about basic demographic data such as age, gender, major of study, life-experience variables such as paid and non-paid work (e.g. charity) experiences, and other socio-economic and family background variables like financial situation, number of siblings and being religious.



(A) Structure of the experiment



(B) Proposer's perspective

(C) Responder's perspective

FIGURE 2.1: The experimental design

2.4.1 Offers and Acceptance Rates

We first analyze subjects' offers. In each of the treatments we do not find a statistically significant (at the 5% level) difference (using a Fligner-Policello robust rank order test) in the offers made by the proposers in groups A and C (that is expected since subjects in these groups play the role of proposers first). The same is true when we compare the pooled offers of groups A and C to group B's offers. Table 2.2 presents the offer data pooled across the three groups within each treatment.

TABLE 2.2: Number of times each offer amount is observed

	£0	£1	£2	£3	£4	£5	£6	£7	Total
Response treatment	2	4	9	80	48	3	4	0	150
Strategy treatment	8	2	25	75	37	2	0	1	150

The offer distributions in the two treatments are different (a Fligner-Policello robust rank order test yields a $p-value$ of 0.007). The average offer is higher under the response method (£3.29) than under the strategy method (£2.95).

Under subjective expected utility, this difference can be driven by either or both the proposer's preferences and beliefs being different in the two treatments. If the proposer believes that the responder is less likely to accept an offer when it is revealed to her before her decision (response method) than when it is not (strategy method), the offer that maximises her monetary earnings will be higher in the response than in the strategy treatment. On the other hand, if the proposer's preferences are less kind to the responder's rejections in the strategy than in the response treatment, that will lead her to make a smaller offer in the former than in the latter treatment. Since we find evidence that the proposer's beliefs are very similar in both treatments, we conclude that the offers in the two treatments are different because the proposer's preferences are different across treatments.

The top two rows of Table 2.3 display subjects' conditional acceptance rates in the responder role in each treatment.

Within each treatment we do not find significant differences in responders' acceptance decisions between groups A and C. We use Fisher's exact test to compare responders' decisions to each of the feasible offers and only find significant differences at the 5% level for the offer of £3 in the response treatment and for the offers of £1 and £6 in the strategy treatment. When we pool the responders' behaviour of groups A and C and compare it to group B's, we find no significant differences between them in the response treatment (only for the offer of £4 do we find a significant difference), but find differences for all offers in the strategy treatment. The acceptance rate of group B is statistically significantly lower than that of the pooled groups A and C for all offers. The differences range from 2.9% (for offer of £4) to 16.9% (for offer of £7).

When we pool responders' behaviour across all three groups within each treatment, we do not find any differences (using a Fisher's exact test and a significance level of 5%) between the treatments' conditional acceptance rates (which we display in the first two rows of Table 2.3).

TABLE 2.3: Conditional acceptance rates: Responder's choices and Proposer's stated beliefs

Offers	£0	£1	£2	£3	£4	£5	£6	£7
Responder's actual choices								
Response trm.	0.100	0.300	0.422	0.911	0.973	1.000	0.900	–
Strategy trm.	0.067	0.305	0.529	0.899	0.973	0.959	0.933	0.893
Proposer's stated beliefs								
Response trm.	0.024	0.164	0.322	0.660	0.869	0.942	0.975	0.989
Strategy trm.	0.033	0.153	0.312	0.619	0.815	0.879	0.896	0.905

We now turn to individual subjects' behaviour as responders. We first check if a subject's strategy as a responder is deterministic, i.e., if she always either accepts or rejects all offers of the same amount. This is a relevant issue because a particular type of deterministic strategies (more specifically, cut-off strategies, i.e.,"reject all offers smaller than x, accept all offers greater than or equal to x") is often used in theoretical models of bargaining. Furthermore, in previous one-shot ultimatum game experiments that used the strategy method, responders' were restricted to playing deterministic strategies. In our study, 89.3% (134) and 52.7% (79) of the subjects always respond in the same way when offered the same amount by different proposers in the response and strategy treatments, respectively.[11]

[11] In the strategy treatment, 70 out of 150 subjects make their accept/reject decisions for the first proposer and then simply press a key to automatically play them in each of the pairings with the other nine proposers. Another 9 subjects who choose not to automatically play the decision for the first pairing at least once, still end up playing it in all pairings. The remaining 71 subjects (play non-deterministic strategies) do not change their strategies very often either: 67.6% (48) of them still automatically copy one of the strategies they play more than 6 times out of 9 times. Interestingly, we find that subjects who have the experience of being a proposer before being a responder (subjects in groups

The amounts most often both accepted and rejected by individual subjects are £2 and £3. Most of the deterministic strategies are cut-off strategies. Overall, 87.3% (131/150) and 49.3% (74/150) of subjects in the response and strategy treatments use cut-off strategies. Given that across the two treatments one third of the responders do not employ a deterministic strategy, we next ask whether they use a cut-off strategy that is deterministic everywhere except in the cut-off (an amount that the individual sometimes accepts and other times rejects, perhaps due to being indifferent), which we call **monotonic**:

Definition 2.1. A responder's strategy is **monotonic** if s/he accepts all offers greater than x, where x is the lowest offer the subject accepts. Otherwise a responder's strategy is **non-monotonic**.

Although, overall a high percentage (82.7%) of responders' strategies are monotonic, the percentage is much higher in the response treatment, 96.0%, than in the strategy treatment, 69.3% (the *p-value of* a Fisher's exact test is 0.000).[12] This suggests that seeing an offer before deciding leads to more systematic responder behaviour. It also suggests that constraining subjects to use cut-off strategies under the strategy method distorts the choices of one third of them.[13]

A and C), use deterministic strategies more often than subjects (in group B) who do not have such experience, 61% vs. 36% (a Fisher's exact test yields a p-value of 0.005).

[12]Among all non-monotonic subjects, half of them reject relatively high offers, while the rest are indecisive about relatively low offers. With respect to those 6 subjects whose strategies are not monotonic in the response treatment, 3 of them reject offers of £5 or higher and the other 3 reject offers of £3 or £4 while accepting lower offers. In the strategy treatment, 27 out of 46 non-monotonic subjects reject offer of £5 or higher at least once. Among the remaining 19 subjects, 7 are non-monotonic only on offers lower than £2, while the others' strategies are not systematic for offers up to £4.

[13]Interestingly, we find that subjects whose responder's strategy is monotonic earn more. The average (across all 10 responses, i.e., before selecting the two responses that determine payment) earnings of the monotonic responders are higher than that of the non-monotonic responders, £3.04 vs. £2.10 in the response treatment and £2.71 vs. £2.36 in the strategy treatment. Kolmogorov-Smirnov tests show significant differences at the 5% level in both treatments.

2.4.2 Stated Beliefs

Table 2.3 also presents the average of proposers' stated beliefs about the responders' probability of acceptance of each feasible offer in each treatment.[14] The average estimates are close in both treatments, differing by less than 6 percentage points for offers up to (and including) £4. A two-sample Kolmogorov-Smirnov test only finds differences at the 5% significance level for offers of £4 and above, which represent only one third of the offers actually made (see Table 2.2). Therefore, given that proposers' beliefs differ little between treatments, and assuming subjective expected utility, the difference in proposers' offers between treatments has to be explained by their preferences being different in the two treatments.

Another thing to notice in Table 2.3 is that at the aggregate level proposers underestimate the probability of acceptance of offers of £5 or less in both treatments.

However, at the individual level we find evidence of a mild projection bias effect. First, the jump in the proposer's beliefs (about the conditional acceptance rates of the responder) for the subjects who use cut-off strategies in the role of responder from the offer £1 below the cut-off to the cut-off amount is much higher than the jump for any other two consecutive offer amounts.[15] Second, we cannot reject the null hypothesis (using a binomial test) that the subject's belief is different from her behaviour as a responder for five out of eight offer amounts for 76% of the subjects in the strategy treatment.[16] Therefore, the subject forms her beliefs about the responder based on her behaviour in that role.

Although we cannot infer subjects' preferences (or, how the Golden Rule influences their offers in the two treatments), we know that if there

[14] As usual (see Costa-Gomes and Weizsäcker (2008)) we note that subjects' stated beliefs are coarse: 81.67% of the stated beliefs are numbers that have one decimal place, while 95.2% have two decimal places, with the second digit either being 0 or 5.

[15] The average jump in the belief is 30.03 percentage points (p.p.), while it is only 16.17 percentage points from the offer £2 below the cut-off to the offer £1 below the cut-off, and 19.91 percentage points from the cut-off offer to the offer £1 above the cut-off.

[16] We cannot perform this test in the response treatment because unlike in the strategy treatment subjects are not offered all feasible amounts ten times.

were no treatment effects, players would maximize their expected monetary earnings equally often and deviate from such maximization in similar ways in both treatments, given that their beliefs are very similar in the response and strategy treatments. To assess this hypothesis we first identify for each subject the offer that maximizes her expected monetary earnings as a proposer given her stated beliefs about the responder's probability of acceptance of the feasible offers.[17] The distributions of such offers (presented in the right column of each panel of Table 2.4 are not different from each other in the two treatments (a Fligner-Policello robust rank order test yields a $p-value$ of 0.96).

Next, we compare a subject's offer with her expected monetary earnings maximizing offer (henceforth *EMEMO*). The second to the fourth column of Table 2.4 present the results ("=", "-", "+" means the offer is "equal to", "smaller than", "larger than" *EMEMO*).[18] By looking at the "Total" row we may conclude that the *EMEMO* accurately predicts half of the subjects' offers in both treatments.

However, the subjects who deviate from the EMEMO exhibit different patterns. While in the response method, subjects deviate from the EMEMO in both directions equally, in the strategy method they tend to go for a lower offer than the EMEMO (a Fisher's exact test for the "+" and "-" categories in the two treatments yields a $p-value$ of 0.044). More concretely, when a subject's *EMEMO* is £4 she makes an offer different from it equally often in both treatments. However, when subjects' *EMEMO*s are £3, they deviate from it differently in the two treatments: in the response treatment, subjects hardly ever choose an offer lower than £3, while they do so one sixth of the time in the other treatment.

These deviations can only partly be explained by subjects' attitude towards risk: only 53.33% $(32/60)$ of the subjects who offer more than their

[17]Our working assumption is that subjects state their beliefs truthfully, even if Costa-Gomes and Weizsäcker's (2008) results suggest that might not always be the case.

[18]There are 16 subjects who state beliefs that have two *EMEMOs*. We assign such cases to the amount they offer, or to largest *EMEMO* amount when they offer more than the *EMEMOs*, or to the lowest *EMEMO* amount when they offer less than the *EMEMOs*. There are 3 subjects whose offers are in between their two *EMEMOs* and we exclude them from Table 2.4. Their behaviour cannot be explained by their risk attitudes.

TABLE 2.4: Offer vs. Expected Monetary Earning Maximizing Offers

Response Treatment

EMEMO/Offer	=	-	+	Total
£0	0	0	0	0
£1	2	0	5	7
£2	3	1	7	11
£3	46	2	19	67
£4	26	31	3	60
£5	0	2	1	3
£6	0	0	0	0
£7	0	0	0	0
Total	77	36	35	148

Strategy Treatment

EMEMO/Offer	=	-	+	Total
£0	1	0	2	3
£1	1	1	4	6
£2	3	0	6	9
£3	44	11	12	67
£4	22	37	0	59
£5	0	4	1	5
£6	0	0	0	0
£7	0	0	0	0
Total	71	53	25	149

EMEMO can be explained by risk aversion, and that 75.28%(67/89) of the subjects who offer less than their EMEMO can be explained by risk loving.[19]

The evidence from this individual level analysis reinforces our view that the main driver for the difference in the distribution of subjects' offers in the two treatments (with the mean also being lower in the strategy treatment) is the difference in subjects' preferences in the two treatments.

[19] Often the risk attitude's explanation fails because the subject makes an offer to which she attaches a probability of acceptance not higher than the one attached to the amount £1 lower than it.

2.5 Role-Reversal Consistency

We now analyse each subject's joint behaviour as a proposer and a responder with the aim of understanding whether most subjects do not treat others in ways that they find not agreeable to themselves, which we call role-reversal consistency. In the context of the ultimatum game we define this concept as follows:

Definition 2.2. A subject is **role-reversal consistent** if she always accepts the offer she makes as a proposer. She is **role-reversal inconsistent** if she rejects that offer at least once. When the subject is never offered the offer she makes, the subject is **role-reversal inconclusive**.

The definition of role-reversal consistency seems to be a close interpretation of the principle "never impose on others what you would not choose for yourself" in the ultimatum game. The rejection of an offer reflects the subject's dislike of being treated that way. Thus, when a subject makes an offer to the others that he himself rejects as a responder, he treats others in a way he himself finds not agreeable, and thereby violates the principle. On the other hand, when a subject accepts an offer equal to her own offer, that suggests that she does not dislike this offer, which, together with the fact that she actually makes this offer to the others, implies that she does not treat others in a way she herself finds not agreeable. Thus, whether or not a subject accepts the offer she makes as proposer is what determines her role-reversal consistency. Note that a subject who rejects an offer higher than the offer she makes will not be role-reversal inconsistent (or violate the Golden Rule), so long as a responder she accepts her own offer. Therefore, a subject's strategy as responder does not need to be monotonic in order for her to be role-reversal consistent.

In our data the large majority of our subjects, 82.01%, is consistent. However, the relative frequency of consistent subjects is higher in the response treatment, 92.97% (out of 128 subjects, as we have to exclude 22 subjects who are not offered the amount they offer as a proposer) than in the strategy treatment, 72.67%.[20] [21]

Thus, it seems that receiving feedback about how others react to the way one treats them and observing how others treat one, helps one to accept that others treat one the way one treated them, or to not treat others in a way that one herself would not like to be treated. This finding has several, not mutually exclusive, interpretations. One is that a subject understands a strategic situation better (and therefore, finds it easier to apply whichever principle guides her behaviour more consistently) when playing a role, if not only she has the experience of playing the opposite role, but also, and more importantly, if both players observe each other's behaviour.[22] Another interpretation is that a subject will ignore principles that she uses to guide her behaviour more often, when she knows that her actions are not immediately observed by her opponents.

The finding that most subjects are consistent is in line with the findings of Güth et al. (1982), despite the fact that the treatment they used to study this issue is markedly different from our study. Güth et al.'s (1982) subjects

[20]A Fisher's exact test for the difference in proportions yields a $p-value$ of 0.0001. The much higher frequency cannot be explained by the fact that in the response treatment subjects who make low offers are less often offered the amount they themselves offer at least once than the subjects who make higher offers. Out of the 22 subjects who are never offered the amount, 1,4,6,5,3 and 3 subjects offered £0, £1, £2, £4, £5 and £6, respectively. Even if we were to assume that subjects who were offered £0, £1 and £2 would themselves reject such offers, a Fisher's exact test for the difference in proportions would yield a $p-value$ of 0.0039., thus rejecting the null hypothesis that role-reversal consistency is the same in the two treatments.

[21]Since subjects in the strategy treatment have to decide whether to accept or reject the amount they offer as proposer 10 times, while in the response treatment they make such decision only as many times as they are offered that same amount by the proposers, one might think that the design makes it harder for subjects to be role-reversal consistent in the strategy treatment. To account for this possibility we compare the consistency rates of the sub-sample of subjects who use deterministic strategies as responders (i.e., make the same decision for all offers of the same amount), finding that the difference between the consistency rates is still statistically significant (100% and 92.4% in the response and strategy treatments, respectively, and a $p-value$ of 0.004). We thank Dirk Engelmann for this suggestion.

[22]This echoes Roth's (1995) interpretation of the behaviour observed in Binmore et al. (1985)'s role-switching two-period alternating offers bargaining experiment. In this experiment subjects who are second-movers in the first game make offers closer to the subgame perfect equilibrium in the the second game as first-movers. Roth's interpretation is that the experience of playing the role of second-mover might help a subject better anticipate the behaviour of someone in that role when playing the role of first-mover.

were asked to simultaneously write down their demands as proposer and responder. They call a subject's demands consistent/conflict/anti-conflict (more recently, Güth and Kliemt (2010) refer to consistency as intra-personal coherence) when their sum is equal to/greater/smaller than the total amount to be divided. Out of their sample of 37 subjects, they found 5 subjects whose demands were in conflict, 15 which were consistent and 17 which were in anti-conflict. Since according to our definition both consistent and anti-conflict demands are role-reversal consistent, in their experiment 86.49% of the subjects were role-reversal consistent.

If we use a notion similar to Güth et al.'s definition of consistency, which we would call **strict consistency** (the lowest offer a subject always accepts as a responder is the one she makes as a proposer, but allow the individual to reject higher offers than it), we find rates of 51.6% and 30.0% (statistically significantly different from each other) in the response and strategy treatments, which are on average identical to Güth et al's rate (40.5% (=15/37)). Note that any subject who is in anti-conflict in Güth et al.'s sense cannot be strictly consistent in our sense.[23]

Finding that most subjects are role-reversal consistent raises two questions: one is which factors, if any, explain why people behave that way; the other is to ask whether consistent and inconsistent subjects perform differently.

We start with the latter. We find that in the response treatment a consistent subject earns an average of £6.38 per pair of proposer/responder decisions, a higher figure than the one, £5.77, earned by inconsistent subjects (£5.93 vs. £4.95, in the strategy treatment).[24] Equally interesting, consistent subjects' stated beliefs are closer to the actual acceptance rates than inconsistent subjects. Indeed, the average of subjects' absolute deviation between the actual and the predicted conditional acceptance rates is

[23]The higher rate of strict consistency in our response treatment might have to do with 35 subjects not receiving offers £1 lower than their own offers. Therefore, we assume that the offer they make is the lowest they would accept. If we exclude these subjects from the analysis, our strict consistency rate is 33.3% (31/93).

[24]K-S tests for the equality of the distributions of earnings of consistent and inconsistent subjects yield $p-values$ of 0.0000 and 0.064 in the strategy and response treatments, respectively.

16.75 percentage points for the consistent subjects and 18.38 for the inconsistent subjects (a K-S test for the equality of the distributions of subjects' average absolute deviation prediction error yields a $p-value$ of 0.043).

TABLE 2.5: Probit regressions on role-reversal consistency (reporting marginal effects)

	1	2
Offer	0.079 (3.69)**	0.070 (3.39)**
Monotonicity	0.256 (3.74)**	0.240 (3.56)**
EMEM	0.078 (1.85)	0.098 (1.66)
Response-method	0.157 (2.14)*	0.159 (2.24)*
Prop.time	0.002 (1.78)	0.003 (1.92)
Resp.time^{rm}	-0.002 (2.40)*	-0.001 (2.22)*
Resp.time^{sm}	-0.001 (1.15)	-0.001 (1.10)
Age		0.014 (1.31)
Gender		-0.010 (0.25)
Quant. major		-0.029 (0.71)
Native Speaker		0.012 (0.21)
Work experience		-0.022 (0.41)
Charity experience		-0.010 (0.24)
Has siblings		0.103 (1.43)
# Rooms at home		-0.073 (1.84)
# Family cars		0.033 (1.32)
# Observations	275	261

Absolute value of z statistics in brackets; * (**) significant at 5% (1%);
The 22 inconclusive subjects and the 3 subjects whose offers lie between their two EMEMOs are excluded from regression 1 and 2;
Those who did not complete the post-experiment questionnaire are excluded from regression 2.

In order to identify the extent to which different aspects of behaviour or innate characteristics influence the likelihood of a subject being role-reversal consistent we use a probit regression (see Table 2.5). We regress a subject's role-reversal consistent behaviour binary variable on her offer as a proposer, on a binary variable of whether her offer coincides with her EMEMO (labeled EMEM in regressions), a binary variable of whether her behaviour as a responder is monotonic, the treatment she took part in, the decision time she took as a proposer and as a responder, and a series of demographic, socio-economic and life-experience related variables.[25]

[25] We treat the responder's decision time differently in the two treatments. In the response treatment it is the time taken to decide on all 10 offers (often some offers are equal

We include the subject's offer in the regression because we expect that the higher the offer a subject makes as a proposer the higher the probability that she will accept it as responder, and therefore the higher the probability that she is role-reversal consistent. We include the binary variable of whether the subject's offer coincides with her EMEMO to test whether behaving strategically as a proposer under the assumptions of risk-neutrality and own-payoff maximization is confounded with role-reversal consistency. Such confounding could arise as a result of projection bias. A subject who as a responder uses a cut-off strategy believes others in that same role behave the same way. Therefore, the subject's EMEMO as a proposer is equal to the cut-off. In this case, and if the subject offers her EMEMO, there is a confounding between behaving strategically and being role-reversal consistent. However, there is no confounding if the subject offers a different amount.

The binary variable of whether the subject's behaviour as a responder is monotonic is included because cut-off strategies play such a huge role in the bargaining literature, and because such behaviour reveals that a subject has clear preferences as to what offers to accept and reject (with the possible exception of the cut-off). We include a dummy variable for the treatment the subject took part in to test which of two effects is stronger: the "hot vs. cold" effect which suggests that subjects might be less role-reversal consistent in the response treatment, as they are likelier to reject (when in the responder's role) the low offer they make as a proposer; or the "experience effect", since subjects play the second role in the response treatment after having observed the behaviour of their opponents in the other role (either how they treated her or whether they accepted how she treated them, depending on which role she had played first), which might make fairness or moral principles like the Golden Rule more salient.

The results in Table 2.5 show that a higher offer, and monotonic behaviour as a responder, have a positive effect on role-reversal consistency.

to each other) received. In the strategy treatment it is time spent entering a strategy, i.e., a conditional decision for each of the eight feasible offers for each proposal averaged across the 10 proposals. If when specifying a strategy to play a proposer the subject simply invokes the strategy she specified for the previous proposer, the decision time spent for that proposal is assumed to be zero.

The response treatment also yields a higher consistency level, which confirms the non-parametric analysis above. The results also confirm that behaving strategically, i.e., choosing the offer that maximizes one's expected monetary earnings does not explain the high rate of role-reversal consistency, and therefore allows us to say that the observed behaviour is not driven by the confounding of these two explanations in our study.

Among the proposer's and the responder's decision time variables (the latter sorted according to treatment) in the regression, the responder's decision time in the response treatment is the only one whose statistical significance is robust to the specification chosen. Its effect is negative, which means that taking longer to decide leads to less consistent behaviour. An intuitive explanation is that the moral principle behind role-reversal consistency is simple and easy to apply. Longer decision times might reflect the individual taking into account a variety of (possibly competing) considerations when deciding, or her use of harder to apply heuristics.

We find that none of the demographic, socio-economic, life-experience, and family related variables affect role-reversal consistency.

2.6 Conclusion

This chapter reports an experiment that tests whether people do not treat others in ways that they themselves would not like to be treated, which we refer to as role-reversal consistency. In the experiment, each subject plays both roles of a modified version of the standard ultimatum game and states her beliefs as a proposer about the responder's behaviour either under the response or the strategy method. In our design each subject when playing as a proposer makes the same offer to multiple responders simultaneously, and when playing as a responder receives offers from multiple proposers, in a way that preserves the one-shot nature of the interactions. This feature improves the comparison of the second mover's behaviour in a two-person two-move game under the response and the strategy methods, because she is asked to make decisions for different actions of the first mover under both methods. Hence, in our design, the main difference between the two methods is whether the second-mover responds to “hypothetical” (in the strategy method) or “actual” (in the response method) actions of the first mover. This differs from the typical previous design comparing both methods in which the second-mover makes one decision given one actual action of her opponent in the response method, but makes decisions for multiple hypothetical actions of her opponent in the strategy method.

Overall, we find that the majority of subjects are role-reversal consistent, a finding that is not driven by subjects behaving strategically as proposers (i.e., choosing the offer that maximizes their expected monetary earnings). We also find that the response method produces a substantially higher level of role-reversal consistency than the strategy method. Furthermore, a larger offer and a monotonic response to others’ offers are more likely to produce role-reversal consistent behaviour. On the other hand, we find that subjects who take longer to decide as responders in the choice-method are less consistent. In addition, we find that role-reversal consistent subjects are better predictors of the behaviour of their opponents, and earn more money than inconsistent subjects.

We also add to the debate on the differences in subjects’ behaviour in the strategy and the response methods under subjective expected utility, by showing that in a two-stage game, the first mover behaves differently in the two treatments not because her beliefs about the second mover are different, but because her preferences are different across treatments.

Finally, the observation that the proportion of role-reversal consistent subjects is high indicates that theoretical models of strategic interactions that do not rely on the usual assumption of common knowledge of rationality, but instead assume that an agent makes a decision considering how she would play the opponent’s role, might have a role to play in predicting both one’s and one’s opponent’s behaviour. Such line of inquiry might be appealing in situations when an agent knows very little about her opponents’ preferences, characteristics, personalities, and so on.

Chapter 3

Expected-Norm Consistency: An Experimental Study of Trust and Trustworthiness[1]

3.1 Introduction

Do you trust the others the way you expect the others to trust you? Are you as trustworthy as you expect the others to be? The first of 7 Principles of Admirable Business Ethics[2] is Be Trustful. Considered as an important component of social capital (Coleman, 1990; Arrow, 1972), trust plays a vital role in people's daily life as well as the functioning of the economy. Expectation is important in people's decision making process, while expectation of social norm enters the process when people are not certain about what kind of norm the others would follow under relative strange circumstances.

Despite the vast literature on trust, the empirical analysis of the relationship between trust and the expectation of norms has been neglected, or

[1]This chapter is based on *Do I play the way I expect others to play my role? Evidence from a expected norm trust game experiment*, a joint work with Miguel Costa-Gomes and Yuan Ju.

[2]See http://sbinformation.about.com/od/bestpractices/a/businessethics.htm. Note that there is also a version of 12 Ethical Principles for Business Executives, http://josephsoninstitute.org/business/blog/2010/12/12-ethical-principles-for-business-executives/

to the least, lacks systematic study. The unanswered question is "To what extent do people follow social norm in a situation like to trust or not to trust?". There are two types of social norms: one is injunctive norm which describes what people should do; the other is descriptive norm which describes what people do. In a strategic situation, people choose their strategies by applying the injunctive norm to themselves and their opponents, while people relates to the descriptive norm by conforming to their perception of it. The existence of the injunctive norm requires people to believe that others would follow the norms. However, in a relatively new situation with stranger opponents, it is difficult to be certain whether their opponents would follow whichever injunctive norm they think is appropriate. Instead, people form expectations about what their opponents would do. While we cannot precisely measure the injunctive norm people have, we can elicit their expected norm and check whether their own behaviour follow such norm.

This chapter aims to develop our understanding of trust in relation to the expectation of norms. More specifically, we do this by experimentally testing whether people's trust and trustworthiness behaviour conform to their own expectation of the norms the others would follow when playing a trust game. The experimental design adopts role-reversal to get subjects to play both roles, in a way that allows us to check whether a subject's behaviour in both roles conforms to their expected norm, or in our terminology, is expected norm consistent.

In a trust game stylised by Berg, Dickhaut and McCabe (1995), one player (the trustor) decides how much out of a certain amount of money to send to his opponent (the trustee). The amount sent is tripled on the way to the trustee who then decides how much out of the tripled amount to return to the trustor. The amount sent is normally considered as a measure of trust, and the amount returned as a measure of trustworthiness.

Given that by the Oxford Dictionary trust is defined as firm belief in the reliability, truth, or ability of someone or something, we know that trust is all about the trustor's expectations or beliefs of the trustee. Therefore, to answer the question of whether people follow their expiated norm when playing the trust game, a crucial issue is to figure out a players' expectation of norm the others would follow. Then in order to see how he actually do, we ask such a player to play the opposite role. By comparing a player's expectation and his actual action, we could determine whether he is expected norm consistent. Moreover, given the two roles of the game, we can separate the expected norm consistent trust from the expected norm consistent trustworthiness, with the former to compare a trustor's actual amount sent with his expectation of the amount being sent by the (new) trustor when he becomes the trustee, and the latter to compare a trustor's expectation of the amount returned by the trustee and the actual amount he returns when he acts as the trustee.

Subjects in our experiment play either the trustor or the trustee first, being unaware that they will play the opposite role later. We use a strategy method for the trustees to elicit a complete strategy profile of how much they would return given different amount they could potentially receive. After their decisions, they are then asked to state their beliefs about the probabilities of each possible amount to be sent or returned by their opponents (incentivised according to the accuracy based on a quadratic scoring rule).

In one treatment, subjects are informed the outcome of the game after stating their beliefs (henceforth, shortened to feedback treatment). In another treatment, such information is withheld from the subjects until the end of the whole session (henceforth shortened to no-feedback treatment). Next, subjects play the opposite role and state their beliefs about their opponents, knowing that they are not facing the same opponents as in the previous game. Both the amount sent and returned are restricted to be whole pounds so that we could use the strategy method for the trustees and elicit subjects beliefs about the probability of each possible amount sent and returned.

Our main findings are as follows.

First, more than half of the subjects are expected norm consistent either in trust or trustworthiness. However, 60.67% of the subjects are expected norm consistent in trust, which is significantly higher than the level of expected norm consistency in trustworthiness at 50.33%.

Second, experience of playing the trustor role has a negative impact on the expected norm consistency in trustworthiness, but no significant effect on the expected norm consistency in trust.

Third, players who are expected norm consistent in both trust and trustworthiness get the highest payoff when play the trust game among themselves.

Fourth, we find that none of the demographic data has a significant impact on the expected norm consistency.

Finally, we contribute to the literature of comparing the trust and trustworthiness behaviour with or without experience and feedback information. Our result shows that experience of playing the opposite role and feedback information has a significantly negative impact on trustee's behaviour, while such impact is not due to a not fulfilled expectation.

3.2 Related Literature

Two distinctive features of our experimental design are the expected norm design and belief elicitation. Given the vast literature of the trust games, we only discuss the experimental work whose design either involves a two-role design or belief elicitation.

In our experiment, we ask the subjects to play both roles, one at a time with different opponents. The order of playing these two roles is varied for different subjects. Such a design allows us to have a clean test of the experience effect of playing the opposite role. Combining with the comparison between the two treatments with or without feedback from their opponents, we could also identify the mix effect of experience and feedback.

In trust games where subjects play both roles, most existing studies have subjects play both roles together (Tu and Bulte, 2010; Tan and Vogel, 2008; Altmann et al., 2008 and Walkowitz et al., 2004) or play the trustor role first, followed by the trustee role, without receiving feedback in between playing the two roles (for example, Garbarino and Slonim, 2009; Chaudhuri and Gangadharan, 2007 and Sapienza et al., 2013 among the existing many trust game studies). Yet, neither of these studies finds any significant difference in subjects' behaviour in either role due to playing both roles of the trust game.

One different result comes from Burks et al. (2003). In Burks et al. (2003)'s design, subjects play trustors first and trustees next, but with different opponents in each role. In some treatments, they know they will play the trustee role next, while in others they don't. There is also a control treatment where subjects would only play one of the roles. Trustors in the treatments where they know they will play the trustee role send significantly less. For trustees, the experience of playing the trustor role first significantly reduce the fraction of money returned to the trustor. Players do not receive feedback in between playing the two games, but they do not use the strategy method for the trustees, which means the trustees can see how much the trustors send them.[3]

Burks et al. (2003)'s interpretation of the golden rule focues on how sympathy and revenge affects the play of either role. We links a person's action in a role closely to her expectations of the others playing the opposite role.

In order to elicit the belief of how the subjects expect the others to treat them, we explicitly ask them to state the probability distribution over all possible actions that could be taken by their opponents. This task is given to both the trustors and trustees.

The novelty of our method of belief elicitation is to elicit the subjects' complete probability distribution among all possible amounts rather than a

[3] It is also possible that the trustees' expectation of the trustors' are not fulfilled, so that they return less to express the dissatisfaction. Also the regressor 'no prior' (no prior knowledge) is not significant in all of their regressions, and as they have suggested the trust level in 'no prior' treatment is lower, which indicates the possibility that the lower level of return is because of lower level of amount being sent. The 'amount received' variable remain marginally significant through out all of their regressions. This also highlights one of our design feature. By using the strategy method and compare the average fraction returned or the total of return, we make sure our test for experience and feedback is conditioned on the amount received.

simple point estimation as it is traditionally done. As Johansson-Stenman et al. (2011) pointed out, rather than a point estimation, subjects generally have a subjective probability distribution over different amount returned. It is generally avoided eliciting the full probability distribution because if the amount sent is relative large, after being tripled, it would be a very confusing and labouring job for the subjects to answer dozens of questions on probabilities. Different from the previous works, we ask subjects to state a probability distribution over every possible amount sent/returned.[4] Subjects are paid according to the accuracy of their predictions using a quadratic scoring rule.

Although belief elicitation is frequently used to determine players' expectations of their opponents' behaviour in trust game, few have done with monetary incentive or for both roles.

In most experiments with the belief elicitation task, the focus has been put on the beliefs of the trustors. Trustors are asked to state how much they expect the others to return to them given their amount sent. Trustors' expected amount to be returned by the trustee is found to be positively correlated with their amount sent. See for example Sapienza et al. (2013), Chaudhuri and Gangadharan (2007), Ashraf et al. (2006), Fehr et al. (2003) and Bolle (1998).

Few experiments of the trust game elicit the beliefs of the trustees about the trustors behaviour. Wilson and Eckel (2006) conduct an incentivised belief elicitation task (asking for a best guess of amount sent) for trustees' and found that for a group of subjects, trustees punish trustors by returning less if their expectation of amount sent is not met. But such result is not robust in other settings, among the experiments that involve eliciting the beliefs of both roles. For example in Tu and Bulte (2010), Eckel and Wilson (2004) and Buchan et al. (2008), they find a significant positive correlation between trustors' belief and their decision, but no relation between the trustee's beliefs and their decisions of return.

[4]The maximum number of possible outcomes is ten. As it turns out, most subjects have no difficulty understanding the task.

However, experimental data from the above mentioned experiments do not allow us to compare one's belief in one role and his own action in the other, for which both the expected norm design and the belief elicitation task is necessary. In Tu and Bulte (2010), subjects play both roles together and state their beliefs for both roles. However, due to the fact that the trustees' behaviour is not their main focus, they do not use strategy method, thus the number of observations, i.e., matched pairs of a trustor's belief given his amount sent and his return behaviour for the same amount sent in the trustee's role would be rather small. Therefore, it is not difficult to see that in order to find out the answer of our main research question in this chapter, it requires a specific design, which we will discuss in details in the following section.

3.3 Experimental Design

A main goal of our study is to examine the links between a subject's decisions in a role and her stated beliefs about how others would play that same role when she plays the opposite role in order to detect evidence for projection bias and check whether behaviour conforms to *expected-norm consistency*. Other goals are: i) to test whether a subject's experience of playing one of the trust game's roles has an effect on her strategy and stated beliefs in the opposite role;[5] ii) whether there is an additional effect from receiving feedback in one role on the strategy and stated beliefs in the opposite role.[6]

Our goals lead us to adapt Berg et al. (1995)'s classical version of the Trust game as follows. First, the trustor has an initial endowment, but, unlike in Berg et al. (1995), the trustee does not.[7] The trustor's and trustee's

[5]Note that in previous studies in which subjects play both roles of the trust game, subjects either play both roles simultaneously or the trustor role first and the trustee's role next. Therefore, prior to our study only the effect of experience as trustor on decisions as a trustee had been studied.

[6]Note that it is not possible to study the effect that the (own) feedback one receives when playing one role has on one's behaviour in the opposite role without having first played the former's role, i.e., without having experience in that role.

[7]This modification has previously used by Glaeser et al. (2000) and Johansson-Stenman et al. (2011).

different endowments might lead players to transfer larger amounts to the player in the opposite role than in the equal endowment's design of Berg et al. (1995), as in such setting transfers reflect both trust and trustworthiness as well as subjects' aversion to payoff inequality (e.g., the trustor has to transfer some money to the trustee to alleviate the extreme initial payoff inequality). However, since we are mostly interested in exploring the links between a subject's decision in a role and her beliefs about how other subjects would play that same role, and examine how they vary across different conditions, not on measuring the absolute level of trust per se, any difference that our design modification has on the latter is irrelevant for our purposes. Moreover, we expect that the difference between the trustor's and trustee's initial endowment lead social norms (such as trust and inequality aversion) to have a larger role on subjects' strategy and stated beliefs in the game. This helps us to test whether in the trust game a subject's behaviour in a role conforms to her expected norm of behaviour of others in that same role in a cleaner way.

Second, the initial endowment given to the trustor, i.e. £3, is substantially smaller, than in Berg et al. (1995) who give her $10. This together with our restriction that players can only transfer whole (i.e., integer) Sterling amounts to each other (Berg et al. (1995) do the same with US dollars) reduces the cardinality of the action space for either player role (in relation to Berg et al.,1995) in a way that makes it feasible to elicit subjects' beliefs over the support of players' action spaces. The whole amount constraint also makes it impossible for the trustor to send half of her initial endowment to the trustee. Furthermore, since the amount received by the trustee is the triple of the amount sent by the trustor, while the amount sent back by the trustor is a one-to-one transfer, final outcomes that equalize players' payoffs are ruled out.[8]

In the experiment subjects play the trust game twice, once in each role, and make their decisions according to the strategy method. Trustors decide how much to send which we denote as $X (X \in \{0, 1, 2, 3\})$. Before being

[8]The sum of the payoffs of the trustor and trustee, given that the trustor can send £0, £1, £2 or £3, is £3, £5, £7 and £9, respectively. The whole amount constraint rules out payoff equalization.

told how much she received from the trustor, trustees choose the amount they would return to the trustor for every possible amount they could receive from her which we label as $Y_X (Y_X \in \{0, 1, 2, \cdots, 3X\})$. We have two treatments: in one treatment, the *Feedback Treatment*, at the end of the game in which they play the first role assigned to them, subjects receive feedback (i.e. are told the decision of their opponent and their own payoff from which they can infer their opponent's payoff), before they play the opposite role in the next game; in the other treatment, the *No-Feedback Treatment*, subjects only receive feedback after they have played both roles.

In both treatments, after playing each role, but before playing the opposite role (and before receiving feedback in the first game they play in the Feedback Treatment), subjects state their beliefs about which action their opponent chose: trustors are asked to state a probability distribution over each possible amount that could be returned to her out of $3X$ (i.e., over $\{0, 1, 2, \cdots, 3X\}$); trustees are asked to state a probability distribution over each possible amount that could be sent to her (i.e. over $\{0, 1, 2, 3\}$).

We use a three-parts-three-groups scheme (identical to the structure shown in Figure 2.1 (A)) to implement an expected norm design that guarantees that subjects play against different opponents when playing the two different roles, in order to suppress any direct induced reciprocity or revenge. More specifically, subjects are divided into three groups. In each part, subjects from two of the groups play the game against each other, with the third group being idle. In part I, group A subjects play trust games against group B subjects, the former as trustors, the latter as trustees. In part II, group A subjects become trustees and play trust games with group C subjects who play as trustors. In part III, group B subjects become trustors and play trust games with group C subjects who become trustees. Thus, subjects in a group only interact once with the subjects of each of the other two groups. Our procedure also varies the order the two roles are played across subjects: some subjects play the trustor role first, while others play the trustee role first.

This design allows us to observe subjects' behaviour (strategies and stated beliefs) in each of the two roles depending on whether they have previous experience of playing the opposite role and on whether in addition they received feedback in the game in which they played that role. Table 3.1 details the level of *Experience* and *Feedback* in the opposite role conditions under which subjects in the different groups play the roles of trustor and trustee.

This allows us to study the effects (if any) that experience in one role has on the chosen strategy (i.e., on the absolute levels of trust and trustworthiness) and stated beliefs in the other role as well as the added effects (if any) that feedback received in that role has on the chosen strategies and stated beliefs in the other role.

A novelty of our design is that we can measure the effects of experience in one role on the behaviour in the other role in both directions, i.e., from the trustor to the trustee, and vice-versa.

TABLE 3.1: Features of the data

Role	Groups	Treatments	Condition	Label
TRUSTOR	A&C	Both	No Exp. & No Feedb.	GACnenf
	B	No Feedb.	Trustee Exp. & No Feedb.	GBenf
	B	Feedb.	Trustee Exp. & Feedb.	GBef
TRUSTEE	B	Both	No Exp. & No Feedb.	GBnenf
	A&C	No Feedb.	Trustor Exp. & No Feedb.	GACenf
	A&C	Feedb.	Trustor Exp. & Feedb.	GACef

We implemented our design as follows. For each of five sessions of each treatment we recruited thirty students from different majors (with the exception of Economics) at the University of York, who were randomly divided into three groups of ten, A, B and C. This gave us 150 subjects per treatment, about two thirds of which are British. Subjects were then told that the session had three independent parts, and that each group would participate in only two parts, and interact with a different group in each part. Subjects received the instructions for each part they participated in only at the start of that part. Therefore, when playing the first of two roles of the trust game they were unaware they would be asked to play the reverse role later. This prevented subjects from playing the two independent games as a whole or balancing their payoffs, at least when playing the first role assigned to them. Subjects interacted anonymously via a z-tree (Fischbacher, 2007) program. Each subject was paid for their decisions in both roles and for the accuracy of their stated beliefs about the behaviour of another subject in each of the roles, as well.

After all subjects have played both roles they completed a questionnaire that elicited their social background, past trusting behaviour and attitudes towards trust and empathy. Subjects received their earnings (which included a show-up fee of £3) in cash at the end of the experiment.

3.4 Data analysis

TABLE 3.2: Trustor's behaviour and Trustee's stated beliefs

Trustor sends	£0	£1	£2	£3	Average
Trustor's behaviour					
Feedback treat.	10.66%	45.33%	30.67%	13.33%	£1.47
No-feedback treat.	12.00%	42.67%	33.33%	12.00%	£1.45
Pooled	11.33%	44.00%	32.00%	12.67%	£1.46
Trustee's stated beliefs					
Feedback treat.	23.26%	37.90%	26.01%	12.83%	£1.28
No-feedback treat.	21.91%	36.90%	26.73%	14.45%	£1.34
Pooled	22.59%	37.40%	26.37%	13.64%	£1.31

In this section, we present the basic data analysis and briefly summarise the behaviour of the trustor and trustee as well as their stated beliefs about the opposite role's behaviour.

3.4.1 The Trust Level

Table 3.2 presents the trustor's behaviour and trustee's stated beliefs. Subjects in the role of trustors send an average of £1.46 to the trustee (i.e.,

TABLE 3.3: Trustee's behaviour and Trustor's stated beliefs

	Groups A&C			Group B		
Send	£1	£2	£3	£1	£2	£3
Actual return						
Feedback	61%	80%	71%	100%	95%	94%
No-Feedback	89%	106%	105%	104%	107%	111%
Pooled	75%	93%	93%	102%	101%	102%
Expected return						
Feedback	122%	129%	115%	101%	115%	99%
No-Feedback	121%	128%	126%	92%	115%	110%
Pooled	122%	129%	121%	97%	115%	103%

about 50% of the trustor's endowment) which is in line with the existing experimental results.[9] However, subjects' stated beliefs as trustees underestimate the amounts they expect to be sent by the trustor, £1.31 (i.e., 43.7% of the endowment).[10] The gap between the two figures arises mainly because the trustee overestimates the probability of the trustor sending nothing.

Neither the trustor's behaviour nor the trustee's stated beliefs react to whether subjects have prior experience of playing the opposite role (i.e., trustee and trustor, respectively), or to whether or not they received feedback when playing that opposite role.

Comparing the amount sent by the trustors between groups[11], we do

[9] Berg et al. (1995) find that the trustor sends an average of 51.6% of her endowment. In addition, they find that giving the trustor information about outcomes of trust game previously played by others does not significantly increase the amount send by her (53.6%). Walkowitz et al. (2006) and Tan and Vogel (2008) find in a study in which each subject plays both roles sequentially, but without receiving feedback in between the two roles, that the amount sent by the trustor is the same regardless whether it is the first or second role played by a subject. Burks et al. (2003) find that subjects who are told they will play the role of the trustee after playing the role of trustor transfer a smaller fraction of their endowment as trustors than when they do not have such prior knowledge, 65.0% vs. 47.3%. In our experiment, when subjects are first assigned a role, they are not told they will play the opposite role next.

[10] This is in line with existing evidence: Tu and Bulte (2010) report that trustees' (monetary unicentivized) beliefs reveal that they expect the trustor to send 39.7% of her endowment; Wilson and Eckel (2006) report that trustees' (monetary point incentivized) beliefs reveal that they expect the trustor to send about 50% of her endowment.

[11] For simplicity of description, we refer to the label given in Table 3.1 when talking about subjects in specific groups.

not observe significant experience effect of playing the trustee on the trustors' behaviour either. GACnenf[12] play as trustors without any experience of playing the opposite role. Their average send level is £1.49. GBnenf, on the other hand, have the experience (but not the feedback information of how much is being sent by their opponents) of playing trustees before playing trustors and their average send level is £1.42. Experienced players generally send less than inexperienced players, but the result is not significant. (The *p-value* of a Fligner-Policello robust rank order test is 0.7548.) Furthermore, GBef have both the experience and the feedback from playing the trustees. However, the send level of GBef and GACnenf is not significantly different from each other (£1.38 vs. £1.49). Experience with feedback does not have any significant effect on the trust level either. (The *p-value* of a Fligner-Policello robust rank order test is 0.5013.)

3.4.2 The Trustworthiness Level

Table 3.3 describes the trustee's behaviour and the trustor's stated beliefs (about the trustee's behaviour). The top panel presents the average amount returned by the trustee to the trustor conditioning on the hypothetical amount sent (not the amount effectively sent, since we use subjects' strategies as trustees for this computation) by the trustor. On average the trustee sends back to the trustor slightly less (91.9%) than the amount that she receives from him, which is in line with the existing experimental evidence.[13] [14]

We find that subjects in the role of trustees who have prior experience of playing the trustor's role and received feedback in that role return 27.6%

[12]Tests of equal distribution for send level between group A and group C players across treatments are not significant. We thus pool the trustor data of group A and group C from both treatments.

[13]10.0%, 28.3% and 36.0% of our subjects return a larger amount than the £1, £2, £3 sent to them, respectively. In Berg et al. (1995), the trustee returns 90.0% of the amount the trustor sent (i.e. before being tripled). However, in Burks et al. (2003), the trustee returns 120.0% of the amount the trustor sent, which might be the result of the trustor sending larger amount to the trustee than in Burks et al. (2003).

[14]Define the amount a subject in the role of trustee returns when receiving £3, £6 and £9 as Y_3, Y_6 and Y_9, respectively. We say that the subject's strategy as a trustee is **monotonic** if $Y_3 \leq Y_6 \leq Y_9$. 95.67% of the subjects' strategies satisfy the inequalities. 62.0% of the subjects strategies satisfy the strict inequality.

and 26.2% less than the subjects who either did not receive such feedback or who had no prior experience of playing the trustor's role, respectively.[15]

GBnenf[16] play the trustees first without any previous experience or feedback. There is no significant difference between group A and group C on the level of return within each treatment, so we pool the data of these two groups in each treatment respectively.

Comparing to GBnenf, GACenf have an experience of playing trustors but not the feedback information of how much is being returned by their opponents. In addition, GACef have the trustor-experience as well as feedback from their opponents. To identify the experience effect on the returning behaviour, we first compare the return level between GBnenf and GACenf. Group A and C players return slightly less (£0.89 vs. £1.02) when £1 is sent but more (£2.12 vs. £2.02, £3.15 vs £3.07, respectively) when £2 or £3 is sent. However, no significant difference can be found in any of the return levels,[17] which suggest that the experience of playing trustor does not have a significant effect on the trustee behaviours.

However, when comparing GBnenf with GACef, the difference in return level becomes highly significant for every amount sent.[18] GBnenf return on average 33.93% of the amount they received, while GACef return only 24.59%, over 9% less. GACef, on top of playing the role of trustors,

[15]In most studies (for example, Tan and Vogel, 2008 and Walkowitz et al. , 2006), the trustee's behaviour is not affected by whichever role they played first. However, Burks et al. (2003) find that subjects who have previously played the role of trustor although not receiving feedback return lower amount in the role of trustees. One possible explanation of why Burks et al. (2003) find a lower level of return is their design. In their control treatment, they replicate the Berg et al. (1995)'s procedure where both the trustor and trustee are given $10 at the beginning. However, it is not clear whether in their two-role treatment, when their subjects play the trustee role, they are given an extra $10 to make the games in these two designs identical. If the players are not given the initial endowment when playing the trustee in the two role treatment as in the control treatment, the trustees are $10 less than the trustors to begin with. It is natural that players return less when they are in a worse off situation then the trustors than when they begin with the same level of wealth. In other words, it may not be the experience but some other structural difference in the game that drives down the return level in Burks et al. (2003) two-role treatment.

[16]Tests of equal distribution for all the return levels by group B players between treatments is not significant. We thus pool the trustee data of group B from both treatments.

[17]The *p-value* of Fligner-Policello robust rank order test is 0.1865, 0.4206 and 0.8603 when £1, £2 and £3 is sent respectively.

[18]The *p-value* of Fligner-Policello robust rank order test is 0.0001, 0.0136 and 0.0147 when £1, £2 and £3 is sent respectively.

also learn about how much their opponents return given the amount they sent. The evidence suggests that trustees with trustor experience and feedback information send significantly less than those who do not. Experience together with feedback information provides a strong drive for trustees to lower their level of trustworthiness.

This shows that feedback from, not simply experience of, playing the opposite role is the main reason for the reduction of the amount returned by the trustee.[19]

One possible explanation for the above result is that learning about how other subjects as trustees behave gives subjects a reference point of how they should behave in that role, and therefore lower the amounts that they return.

The bottom panel of Table 3.3 describes the trustor's stated beliefs about the amount that the trustee will return to her. As explained earlier, the trustor states a probability distribution over the different amounts that can be returned given the amount that she sent (rather than stating a different probability distribution over the feasible outcomes for each of the amounts that she could send).

Subjects' beliefs revealed that they as trustors expect to receive on average (across treatments) 117.5% of the amount sent to the trustee. Having prior experience as trustee lowers the fraction of the amount sent to the trustee that the trustor expects to be returned, 104.3% vs. 123.9% (for the subjects who do not have experience). Receiving feedback in the role of a trustee has no effect on the subsequent beliefs that subjects state in the role of trustor.

Overall the data shows, that the trustor's stated beliefs overestimate (117.5% versus the 91.9% actually returned by the trustee) the amount

[19] In our design, we cannot explicitly identify the effect on behaviour of receiving feedback from playing the opposite role, since it is necessary for players to experience playing that role in order to get that feedback. Nevertheless, by comparing the fraction of the amount received that is returned between the GACef and GACenf subjects, we find that the feedback variable is what explains the change in behaviour; GACef subjects send about 9% less than GACenf subjects.

returned by the trustee, and this is true regardless of whether or not subjects have have prior experienced as a trustee.[20] However, experience as a trustee moves the subject's beliefs as a trustor somehow closer to the actual behaviour of the trustee.[21]

3.5 Expected-Norm Consistency

In this section, we present our main findings. First, we investigate, separately for each role, the relationship between a subject's behaviour in one role and the beliefs she states about how others would play that role when she plays the opposite role, which we call *expected norm consistency*. We then examine the implications of behaviour that is expected norm consistent in both roles and expected profit seeking behaviour in the role of trustor. Next, we use simulations to determine the benefits and costs of being expected norm consistent when playing subjects who are also consistent vs. subjects who are not. Last, we examine which variables explain subjects' expected norm consistent behaviour.

For trustors who send X, we use trustor's stated beliefs to compute the amount that the trustor expects to be returned and refer to it as y_X.[22] We use trustee's stated belief to compute the amount the trustee expects the trustor to send and refer to it as x.

We now examine the relationship between a subject's behaviour in one role and the beliefs she states about how others would play that same role when she plays the opposite role. We check whether a subject in a role behaves in a way that is consistent with her expected norm of how others would behave in that same role, which we call Expected-Norm Consistent

[20] This goes against the underestimation implied by trustrors' (monetary unincentivized) beliefs in Tu and Bulte (2010) (95.1%) and by trustors' point estimated monetary incentivised beliefs for the different subgroups who played quite different trust games from ours in Garbarino and Slonim (2009) (range between 70.0% and 74.8%).

[21] Similar effects have been observed in the ultimatum game where playing the responder role first helps the proposer better predict the responder's behaviour.

[22] Subjects who send zero are not asked to complete this task, since the only possible amount to be returned is 0, i.e. for them $y_X = 0$.

behaviour. We apply this notion of consistency separately for trust- and trustworthiness-related behaviour.

Definition 3.1. A subject's **trust** behaviour is **expected-norm consistent** if the amount she sends, as a trustor, X, is greater than or equal to the amount she expects to receive from another subject playing the trustor role when she plays the role of a trustee, x, (i.e., if $X \geq x$).

Definition 3.2. A subject's **trustworthiness** behaviour is **expected-norm consistent** if the amount she returns when she receives $3X$, Y_X is greater than or equal to the amount she expects to be returned when she sends X, y_X, (i.e. $Y_X \geq y_X$).

In our data the level of trust consistency is much higher than the trustworthiness consistency, 60.7% (182/300) vs. 50.3% (151/300).[23]

Subjects' coarse beliefs[24]might have a negative effect on the consistency rates on either role. For some inconsistent subjects, their behaviour in one role and the expected value of the beliefs they state about how others would behave in the same role differs by less than £0.1. If we were to ascribe these small differences (i.e., $<$ £0.1) to the effect of subjects' stated beliefs being coarse and classify subjects for whom the differences are small as consistent, the consistency rates go up by fewer than five percentage points in either role. This difference is all driven by the feedback treatment, where the consistency rates are 62.0% (93/150) vs. 45.3% (68/150).[25]

An explanation for the feedback treatment's difference is that experience and feedback lower the level of trustworthiness consistency (especially for groups A and C), but has no significant effect on the level of trust consistency. [26]

[23]These figures assume that the subjects who send £0 are trustworthiness consistent. When we exclude them from the analysis because they were not asked to state their beliefs, the level of trustworthiness consistency is 44.0% (117/266).

[24]We look at the last digit of subjects' stated beliefs and confirm that they are coarse. In total, subjects in our experiment stated 2780 beliefs in percentage number and 96.25% of them have two decimal places with the second digit being 0 or 5. 74.67% of the stated beliefs are either 0 or have only one decimal place. (18.81% of the stated beliefs are 0.)

[25]In the no-feedback treatment the difference is small, 59.3% (89/150) vs. 55.3% (83/150).

[26]An alternative explanation is that 44.7% of the trustors (those who send £2 and £3) state beliefs about a larger number of outcomes (7 or 9 outcomes) than the trustees do

Now we look at the effect of feedback and the order effect of choosing an action in one role and stating the beliefs in the opposite role on the consistency of trust and trustworthiness. There is no feedback effect or an order effect (of choosing an action in one role and stating the beliefs in the opposite role) on the trustor-consistency, but such effects are significant on the trustee-consistency. The proportion of trustee-consistency subjects is much higher in group B than that in groups A and C (64.0% vs. 43.5%, across both treatments, and the $p-value$ of Fisher's exact test is 0.001). Such difference is mainly due to the difference in the feedback treatment (where the proportions are 64.0% vs. 36.0%, the $p-value$ of Fisher's exact test is 0.002), while in the no-feedback treatment, such difference less prominent (64.0% vs. 51.0%, the $p-value$ of Fisher's exact test is 0.164).

Among the four different groups, GACef, GBef, GACenf and GBenf, GACef has the lowest trustworthiness consistency level. There is a significant difference in the trustworthiness consistency level between subjects in GACef and GACenf (the $p-value$ of Fisher's exact test is 0.046). We also observe significant difference between GACef and GBef. The difference between each of the above pairs is whether subjects get feedback information before playing the trustee's role. Our conclusion is that the feedback information about trustees behaviour has a negative impact on the trustee-consistency level.

There is no significant correlation between a subject's trustor and trustworthiness consistency ($\phi = 0.008$). This means that the percentage of subjects who are expected norm consistent in both roles is close to the product of the consistency rates in the two different roles, i.e., 30.3%. The right column of Table 3.4 reports the number of subjects across the two treatments who are consistent in either role, in both role and in neither role.

(always 4 outcomes). The trustworthiness consistency level for subjects who send £1 is 36.4%. Although they also report probability distributions over 4 possible outcomes as the trustees do, their trustworthiness consistency level is still much lower than their own trust consistency level (41.7%). This is evidence that this explanation cannot account for the difference between the trust and trustworthiness consistency levels.

TABLE 3.4: Number of Consistent and Expected Profit Seeking Subjects

Consistent in role	Feedback Treat.			No-Feedback Treat.			Total
	EPS	Not-EPS	Sum	EPS	Not-EPS	Sum	
Neither	27	2	29	27	2	29	58
Trustor	41	12	53	26	12	38	91
Trustee	23	5	28	26	6	32	60
Both	22	18	40	30	21	51	91
Total	113	37	150	109	41	150	300

3.5.1 Trustor's Expected Profit Seeking behaviour

We say that a subject as a trustor is an *expected profit seeker* (EPS) if the amount she expects to be returned to her (y_X) is greater than or equal to the amount she sends to the trustee (X), i.e. $y_X \geq X$.[27] In our data 74.0% (75.3% in the feedback treatment, and 72.7% in the no-feedback treatment) of the subjects behave in this fashion.[28]

We find that having received feedback from playing the trustee's role has no effect on a subject's expected seeking behaviour as a trustor.[29] As discussed earlier subjects who have the experience of playing the trustee's role before the trustor's role (group B) expect lower amounts (which is indeed more accurate) to be returned by the trustee than subjects who do not have such experience (groups A and C). However, since such subjects do not lower the amounts they send as trustors, they end up being expected

[27]In our design, we cannot determine whether a subject's decision as a trustor maximises her expected monetary earnings since we do not the elicit beliefs about how much money the trustee would return for every possible amount that the trustor could send. Interestingly, no design of the experimental studies so far can answer this question.

[28]In line with the expectation interpretation of trust see Gambetta (2000) and James (2002) for general discussion of trust and expectation. Studies that elicit trustors' expectations on the possible amount returned, for example, Garbarino and Slonim (2009), Chaudhuri and Gangadharan (2007), Sapienza et al. (2013), Tu and Bulte (2010), all find a positive correlation between trustors' actual amount sent and expected amount returned. The behaviour of trust could be partly explained by the expectation of others' trustworthiness.

[29]There is no significant difference in the proportion of EPS subjects between the GACef (81%) and GACenf (77%) (*p=0.603* for Fisher's exact test) or between GBef (64%) and GBenf (64%)(*p=1.000* for Fisher's exact test).

profit seekers (statistically significantly) less often than group A and C subjects, 64.0% vs. 79.0%.[30]

For a subject who is both trust and trustworthiness expected norm consistent and is also an expected profit seeker as a trustor, the following inequalities hold, $Y_X \geq y_X \geq X \geq x$, where the middle inequality imposes expected profit seeking behaviour in the role of trustor. Note that these inequalities imply $Y_X \geq X$, i.e., that the subject in the role of a trustee when sent the amount she sends as a trustor, must return no less than that amount. In such case, her opponent will profit from her behaviour. We find these inequalities only hold for about 1/6th of the subjects.

3.5.2 The Benefits and Costs of Expected-Norm Consistency

In this part we look at whether it is profitable to be expected norm consistent when playing against different opponents.

We first summarise the payoff data from the experiment. Consistent trustors send significantly more than inconsistent trustors (£1.91 vs. £0.77, *p-value* for a Fligner-Policello robust rank order test is less than 0.0001). Meanwhile, consistent trustees return significantly more in total for all possible amounts sent than inconsistent trustees: £7.74 (excluding all subjects who send £0 as trustors, because these are not asked for their beliefs; £6.64 if including those) vs. £4.60; *p-value* for a Fligner-Policello robust rank order test is less than 0.0001.

Based on the basic characterisation, there is a substantial number of inconsistent subjects. When the proportion of inconsistent subjects is relatively big, subjects risk their payoffs by being consistent, because the inconsistent subjects send less as trustors and return less as trustees than the consistent subjects. We analyse this by first looking as the real payoff data and then the simulation results.

[30]Fisher's exact test result show that the hypothesis for equal proportion of EPS subjects in GACnenf and GB (pooled data of GBenf and GBef) is rejected at *p=0.008*.

TABLE 3.5: Simulated result for payoffs
(between consistent/inconsistent trustors and trustees)

		trustee	
		Con	Incon
trustor	Con	3.24, 3.57	2.54, 4.27
	Incon	3.03, 1.51	2.75, 1.79

TABLE 3.6: Simulated payoffs for different combinations of expected norm consistency at subject level

Combinations	Consistent in Trust	YES	YES	NO	NO
	Consistent in Trustworthiness	YES	NO	YES	NO
Payoffs	as trustor	3.63	2.39	2.95	2.80
	as trustee	3.38	4.24	0.96	2.41
	Total	7.01	6.63	3.91	5.21

Through out all the sessions and treatments (after random pairing and random selection of the pairing) , consistent trustors earn on average £2.80, while inconsistent trustors earn slightly more at an average of £2.86.[31] Consistent trustees earn on average £2.66 which is significantly ($p = 0.0081$ for a Fligner-Policello robust rank order test) less than the £3.22 earned by the inconsistent trustees.

It is worth finding out what is the earning for the pair when they are both consistent in two roles comparing with those who are both inconsistent in both roles. Consistent pairs earn on average £1.25 more than inconsistent pairs (£5.50 vs. £6.75). Given that the trust consistent subjects send more than trust inconsistent subjects, it is key to have a larger pie to divide between both players.

Next, we analyse the payoff based on a large simulated sample in order to see how being consistent pays. We randomly draw (with immediate

[31]The difference is not significant with *p-value* of a Fligner-Policello robust rank order test equals to 0.7265. The difference when they play against consistent or inconsistent trustees is not big with the highest average of £3, when inconsistent trustors play against consistent trustees, and lowest at £2.62, when consistent trustors play against inconsistent trustees.

replacement) 10,000 subjects' play as trustors and trustees from the consistent and inconsistent categories and match them to calculate each subject's payoff in each role.

First of all, the best scenario for trustors (with average payoff of £3.24) is to be consistent as trustor and play against inconsistent trustee. On the other hand, for trustees, they get the highest payoff (£4.27) when they are inconsistent as a trustee, but their opponents are consistent as trustor. The lowest payoff in each role is achieved by being consistent and playing against inconsistent opponent.[32]

In Table 3.6, we report the simulated payoffs when all players are of the same kind and they play among themselves.[33] Interpreting along the line of Kant's "categorical imperative" , it means that a player's expected norm consistency principle is adopted as a maxim by the population, but in relative terms, that is each player in the population treat other people according to one's own expected norm.

We randomly draw 30 subjects from certain consistent categories and divide them into three groups. Subjects from each group play the game once in each role against one of the other two groups. The payoff of a subject is the sum of payoffs in playing both roles. Payoffs of 10 subjects from one of the three groups are recorded as 10 observations. This procedure is repeated 1000 times to generate 10,000 observations. Comparing the average of these 10,000 observations from different consistency categories gives the following results.

The trustor-consistent and trustee-consistent population comes first, they could earn on average £7.01 for playing the two roles. Next comes the trustor-consistent and trustee-inconsistent subjects, their average earning is

[32] Analysis from a subject level confirms the above results. When we look at a subjects' consistency in both roles as a whole when they play the two roles, the highest earning subjects are consistent as trustor but inconsistent as trustee while their opponent are consistent as both trustor and trustee. They could expect to earn on average £8.23. Subject who are consistent in trustor and trustee get the lowest payoff when they face opponents who are inconsistent as trustor and consistent as trustee. They earn on average £3.45.

[33] In Table 3.5, the consistent trustor is not necessarily trustworthiness consistent(inconsistent), while in Table 3.6, subjects are characterised according to both trust and trustworthiness consistency. This means that the pools from which the strategies are randomly drawn is different.

£6.63, which is not significantly lower than the top population.The both inconsistent subjects population comes the third with the average earning of £5.21; it is significantly lower than the top two populations. The worst situation as we have mentioned is when the subjects are all trustor-inconsistent and trustee-consistent. They earn on average only £3.91.

Subjects who send more are more likely to be consistent as trustor, thus creating a bigger pie between the pair. Inconsistent trustees generally return less thus keeping more for themselves. The general welfare of the population largely depends on the amount sent by the trustors, which is why trust is considered as a social capital. Trustworthiness is important in determining the distribution of the total welfare. When the environment is kind, that is when subjects are consistent as trustor and trustee, their payoffs are the highest.

3.5.3 Regression Analysis

In this section, we present parametric analysis to identify which variables influence the amount sent by the trustor, the sum of the amounts returned of the trustee's strategy ($\sum Return$) and the trustor's and trustee's Expected-Norm consistency.

We start with the first two regressions whose results we present in Table 3.7. We include the variables Experience and Feedback to check whether experience of playing the opposite role and receiving feedback in that role influence either the amount sent by the trustor and/or the sum of the amounts returned of the trustee's strategy. We also include variables that relate to the beliefs that a subject states both as a trustor and as a trustee. The variable Expected Sent Amount (ESA) by Trustor is calculated based on the probability distribution a subject states as a trustee over the possible amounts sent by the trustor. The variable Expected Returned Fraction of Amount Sent (ERF) is the ratio between a trustor's expected amount returned by the trustee and the trustor's amount sent. This variable varies between 0, when the trustor expects nothing to be returned, and 3, when she expects the whole tripled amount to be returned. We also include

TABLE 3.7: Ordered Probit Regression on the Trustor's Transfer and Tobit Regression on the Trustee's Sum of Amounts Returned ($\sum Return$)

	(1) Trustor Transfer	(2) Trustee $\sum Return$
Experience	0.064 (0.43)	-0.605 (1.44)
Feedback	0.026 (0.19)	-1.169 (3.05)**
Expected Sent Amount	0.916 (5.74)**	2.706 (6.66)**
Expected Returned Fraction	1.04 (6.17)**	1.900 (4.28)**
Decision Time	0.004 (2.66)**	0.006 (2.38)*
Male	0.074 (0.51)	0.114 (0.26)
Quant. Major	-0.040 (0.28)	0.250 (0.59)
British	0.374 (2.55)*	-0.233 (0.55)
Trust People	-0.012 (0.14)	-0.335 (1.27)
Others' Feelings	0.073 (0.76)	0.309 (1.18)
Take Advantage	0.039 (0.44)	0.166 (0.64)
Helpful People	0.089 (1.17)	-0.083 (0.37)
# Observations	288	288

* (**) significant at 5% (1%);
Absolute value of z statistics (regression 1) or t statistics (regression 2)in brackets;
Subjects who did not complete the questionnaire are excluded from the regressions.

TABLE 3.8: Probit regression on Trustor's and Trustee's Expected-Norm Consistency

	(3) Trustor		(4) Trustee
Experience	-0.134(1.81)	Experience	-0.184 (2.81)**
Feedback	0.005 (0.07)	Feedback	-0.064 (1.01)
Amount Sent	0.608 (8.69)**	$\sum Return$	0.045 (4.27)**
Decision Time	0.000 (0.47)	Decision Time	0.000 (0.25)
Male	-0.028 (0.39)	Male	0.143 (2.11)*
Quant. Major	-0.057 (0.77)	Quant. Major	-0.100 (1.48)
British	0.090 (1.30)	British	0.032 (0.46)
Trust People	0.019 (0.40)	Trust People	-0.070 (1.68)
Others' Feelings	-0.026 (0.53)	Others' Feelings	-0.011 (0.26)
Take Advantage	-0.38 (0.84)	Take Advantage	0.010 (0.26)
Helpful People	-0.067 (1.72)	Helpful People	0.021 (0.61)
Obs	288	Obs	288

Coefficients are marginal effects; * (**) significant at 5% (1%);
Absolute value of z statistics in brackets;
Subjects who did not complete the questionnaire are excluded from the regressions.

the Decision Time (the amount of time a subject spent on making the decision that appears as the dependent variable in the regression), a series of demographic variables and answers to social attitude questions.[34]

Regression (1) shows that experience in the role of trustee and receiving feedback when playing that role do not influence the amount the trustor sends. This confirms the earlier non-parametric analysis. As in the previous literature, we find that trustor's expectation about the amount returned by the trustee has a positive effect on the amount that she sends. Regression (1) also shows that the amount a trustor sends is greater the larger the amount she expects other subjects to send her if she were the trustee. This is a novel finding of our study. Decision Time is also highly significant but with a small effect. The longer the trustor takes to decide the more she sends.

Regression (2) shows that receiving feedback when playing the role of trustor influences the sum of the amounts the trustee returns as in the earlier non-parametric analysis. Another novel finding of our study is that the sum of the amounts a trustee returns is greater the larger the amount she expects others to send her. We also find that trustee's expectation about the amount returned by the others if she were a trustor have a positive effect on the sum

[34]The demographic variables includes: Male, which takes value 1 if a subject is a male and 0 otherwise; Quant. Major, which takes value of 1 if a subjects study a quantitative major such as Mathematics, Physics, Chemistry and Economics, and 0 otherwise; British, which takes value 1 if a subject's nationality is UK and 0 otherwise.

The variables that describe the answers to the attitudinal questions are the following. Trust People, which represents the answer to the question "Would you say that most people can be trusted?", takes value of 1 if a subject choose the answer "Most people can be trusted"; -1 for the answer "Most people cannot be trusted"; 0 for the answer "I don't know". Others' feelings, which represents the answer to the question "In a situation of conflict, to what extent does the thought of how others will feel about your decision affect it?", takes value of 2 if a subject answers "It significantly affects my decision"; 1 if a subject choose the answer "It has only a small effect on my decision"; -1 for the answer "It does not affect my decision at all"; 0 for the answer "I don't know". Take Advantage, which represents the answer to the question "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?", takes value of 1 if a subject choose the answer "They would try to be fair"; -1 for the answer "They would try to take advantage"; 0 for the answer "I don't know". Helpful People, which represents the answer to the question "Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?", takes value of 1 if a subject choose the answer "They try to be helpful"; -1 for the answer "They mostly just look out for themselves"; 0 for the answer "I don't know".

of the amounts she returns. The longer the trustee takes to decide the more she returns.

All demographic variables and the answers to all the social attitude questions are not significant in regression (1) and (2), except that British subjects send a higher amount as trustors.

To further explore which factors influence Expected-Norm Consistency, we run separate probit regression for each role. See Table 3.8. We use the same explanatory variables as in the previous two regressions, but drop the two expectation variables and add the amount sent by the trustor in the trustor's expected-norm consistency regression and add the sum of the amounts returned by the trustee in the trustee's expected-norm consistency regression.

The amount sent is only statistically significant variable in regression (3). The larger the amount the subject sends as a trustor, the more likely she is expected-norm consistent in the role of trustor. Regression (4) shows that the experience of playing the role of trustor has a positive effect on the expected-norm consistency of trustee. The larger the sum of the amounts returned, the higher probability for the subject to be expected-norm consistent as a trustee.

None of the demographic variables and the answers to all the social attitude questions are significant in regression (3) and (4), except that male subjects are more likely to be expected-norm consistent trustees.

3.6 Conclusion

One important lesson to learn from the previous studies of the trust game is that trust and trustworthiness are very sensitive to the experimental environment. The expectation of what other people would do have an important effect on how people behave. The background of subjects and what information they receive before and during the game can dramatically change the results.

The study of the Expected Norm in strategic interactions is relatively new comparing to the extensive literature on the trust game. Here we find ourselves in examining a complex relationship between people's expectation of how others would do in the opposite role and their own behaviour in that role. We interpret people's expectation of what others do as the Expected Norm. The question is whether people's own behaviour conforms to their expected norm.

We find that more than half of the subjects are consistent with their expected norm, while the level is higher when they are in the trustor's role than in the trustee's role. Trustees, as the second mover, are more likely to violate their expected norm.

Second, experience of playing in the trustor's role has a negative effect on consistency in the trustee's role. The experience of playing the game as a trustor significantly reduces the level of being consistent as a trustee, while whether getting feedback or not from playing the trustor's role does not have a further effect on the consistency level.

Third, consistent trustors and trustees get the highest payoff overall when they play the game against each other. The result also sends out the positive message that in a society with consistent trustors and trustees, the overall welfare is higher than in other cases.

Chapter 4

Let Them Vote!–An Experimental Study of the Public Goods Game[1]

4.1 Introduction

The provision of public goods relies on cooperation among agents who face a conflict between self-interest and social welfare. Neoclassical economic theory and game theory predict that such dilemma typically leads to under-provision of the public goods and loss of efficiency. Experimental studies, however, have found that people do voluntarily contribute to public goods. The average contribution rates range from 40% to 60% (of the initial endowment), and they typically decline over time as the interaction is repeated (Ledyard, 1995).

Both theoretical and experimental research have expanded tremendously over the past decades in the direction of identifying possible ways to solve the free-riding problem and to improve efficiency. One possible way around this "classical conundrum" (as Bowles, 2006, prefers to call the social

[1] This chapter is based on the joint work with Yukihiko Funaki and Róbert F. Veszteg, *Public-goods game with endogenous institution formation: Experimental evidence on the effect of the voting rule*.

dilemma) is to create and rely on institutions that enforce high contribution levels. Most of the theoretical analysis is set in the context of the international environmental agreements (IEAs) negotiation, and in the proposed models, participants first decide whether or not to participate in an institution which forces its members to contribute to the public goods at a level decided collectively. Non-members can decide how much to contribute on their own. Theoretical results suggest that there is no substantial efficiency gain from this mechanism (Barrett, 1994; Hoel, 1992; Carraro and Siniscalco, 1993). The reason is that if an institution was to maximize the joint benefits of its members, it would require the highest possible level of contribution to public-goods provision from its members. Therefore, only a relatively small institution would be formed on a voluntary basis, and participants would have a strong incentive to free-ride.

These pessimistic conclusions seem to have also been reflected in real-life IEAs negotiations among countries. Agreement is typically hampered by the facts that too many countries are trying to agree on a too ambitious contribution level. More often than not, the negotiation breaks down without significantly improvement in efficiency.

Contrary to the pessimism from theoretical models, experimental results show that people are willing to form institutions and provide public goods if the institution is sufficiently large. Consequently, there does exist a real chance for significant improvement on the efficiency level as compared to the non-cooperative status-quo (Kosfeld et al., 2009, Burger and Kolstad, 2009).

In this paper, we focus on a mechanism of voluntary institution formation with a relatively large group of participants. We also aim to find out how the institution-formation mechanism works under different decision protocols, so we analyse the unanimity voting rule or the plurality/majority voting rule. Even if Kosfeld et al. (2009) report promising experimental results where the implemented institution would often include all participants (4 in each group), our conjecture is that in a large group it is more difficult to form a grand coalition. Moreover, comparing to the plurality rule, the unanimity rule is more demanding at the implementation stage, which makes it less likely for the initiated institution to be implemented. One simple way to increase the chances of reaching an agreement is to replace the unanimity voting rule with the plurality/majority rule. However, that could come at the price of a relatively low participation and initiation rate, and consequently, in a smaller institution.

Experimental studies of institution formation are relatively new, few and inconclusive. Beside the rather optimistic message sent by Kosfeld et al. (2009), there is also experimental evidence suggesting that allowing institution members to propose a binding minimum contribution level generates only a marginal efficiency gain. However, if joining the institution is not a voluntary decision but it is forced exogenously on all participants, the efficiency gain can be substantial (Dannenberg et al., 2010; Dannenberg et al., 2014). On the other hand, and similarly to our results, changing the threshold of the number of institution members required to determine the binding minimum contribution level (from simple majority to $\frac{3}{4}$) does not change the overall efficiency level. It simply shows a trade-off between the participation rate and contribution level inside the institution (Dannenberg, 2014).

Our experimental design is based on the one presented and analysed by Kosfeld et al. (2009). Participants first voluntarily decide whether to join an institution. If more than one person decides to do so, a collective decision has to be made about the contribution level that the institution will impose on all of its members. Different from Kosfeld et al. (2009), our design uses the plurality/majority voting rule when members decide whether the institution should be formed and if yes with which required contribution level. Also, in our main treatment members, in order to balance participation rate and contribution level (i.e., successful institution formation), we allowed participants to vote for three different possibilities: to dissolve the institution, to form the institution and to require a 50% contribution rate, and to form the institution and to require a 100% contribution rate. This choice has been inspired by some theoretical analysis of trade-off between the participation rate and the required contribution level inside the institution (Barrett, 2002; Finus and Maus, 2008). The idea is that by lowering the required contribution level an institution might be able to attract more participants to join, which then would lead to a higher efficiency level.

In our experimental study, we find that, as compared to the voluntary institution-formation mechanism with unanimity rule, the plurality/-majority rule induces a lower participation rate, but makes it easier to reach an agreement and to form an institution. Interestingly these opposing effects cancel each other out, which makes it impossible to rank the two voting rules by contribution levels or efficiency. As for the observed institution size, in line with Burger and Kolstad (2009) who show that the grand coalition is rarely formed in large groups, we argue that it is closer to the minimum efficient institution size than to the size of the grand coalition. Meanwhile, note that comparison across papers is rather difficult, because for example Dannenberg (2014) does not use the “traditional” voluntary-contribution mechanism, but relies on a model from the coalition-formation literature (Carraro and Siniscalco 1993, Barrett 1994). Also, in her case, the institution determines a binding *minimum* contribution level, not the contribution level. Nevertheless, the main message is clear, which says a less demanding voting rule combining with less stringent contribution requirements in the institution does not affect the overall efficiency but only results in a tradeoff between the institution size and the implementation rate.

Besides the trade-off of institution formation, the change of mechanism also leads to a significant impact on the contribution levels of players outside the institutions. While under the unanimity rule, players outside the institution (or all the players when no institution is formed) normally choose to free-ride, with the plurality rule, those players still contribute a considerable amount of their endowments. We also show that experience in general, whether playing the game within a same or a different group significantly lower the efficiency level.

4.2 Theoretical Predictions

Our experimental design is based on a simple linear public-goods game in which a group of $n(\geq 2)$ people are required to decide individually how much money they contribute to a public good and how much money they keep for themselves from an initial endowment of $w(>0)$ monetary units.

Given the complete list of contributions $(g_1, g_2, \cdots, g_n)$, agent i's monetary payoff can be written as

$$\pi_i(g_1, \cdots, g_n) = w - g_i + a \sum_{j=1}^{n} g_j, \tag{4.1}$$

where a is the marginal per capita return (MPCR) from contributing to the public good. Note that whenever $a < 1$, contributing zero to the public good is a dominant strategy for each player. Also, as long as $\frac{1}{n} < a$, the group would do better if everybody contributed the entire initial endowment. With other words, for $\frac{1}{n} < a < 1$, the above game has a unique Nash equilibrium in dominant strategies where each player contributes zero, while the welfare-maximizing strategy profile is when everybody contributes the entire initial endowment.

The public-goods game with institution formation is a multi-stage game based on the previously described game. People first express their willingness to form an institution that later in the contribution stage will enforce the jointly-determined contribution level.

- *Participation stage:* Agents simultaneously and independently announce whether they are willing to join an institution which, in the last stage, is going to force its member to make a certain level of contribution (to be determined in the second stage). Agents who choose to join are called members; those who choose not to join are called non-members.
- *Implementation stage:* Members are informed about the total number of members, and they are asked to vote simultaneously in order to decide the contribution level that the institution is going to enforce

in the final stage. Note that members will not be allowed to deviate (positively or negatively) from the chosen contribution level. There are three possibilities to be considered in the institution:

1. Project 0: the institution is dissolved,
2. Project $\frac{1}{2}$: each member contributes half of the initial endowments, and
3. Project 1: each member contributes the entire initial endowment.

The decision in the institution is based on the plurality rule, that is the project that receives the most number of votes is going to be implemented. All members must comply with the decision which in case of projects $\frac{1}{2}$ and 1 is costly. Costs are shared equally among the members of the institution.

- *Contribution stage:* If the institution is not dissolved in the implementation stage, all its members must make a contribution according to the chosen project. Non-members, after being informed of the size of the institution, decide individually how much to contribute to the public good. If the institution is not implemented (i.e., it is dissolved), all agents decide their contribution individually and simultaneously.

Note that this game is essentially identical to the one introduced and studied by Kosfeld et al. (2009). The only difference is that in our design participants use the plurality rule (instead of unanimity) to reach a decision in the institution and they have three projects to choose from. Also the game-theoretical analysis presented below follows closely the one by Kosfeld et al. (2009).

In order to write agent i's final payoff π_i, let S be the set of players who are members of the institution, s the size of the institution ($s = |S|$), and $c \geq 0$ the cost of enforcing project $\frac{1}{2}$ or project 1. If project 0 is chosen, we say that the institution is not implemented and consider S to be empty, so $s = 0$.

If $S \neq \emptyset$, i.e. the institution is implemented, then agent i's payoff is given by

$$\pi_i = \begin{cases} w - g_i + a\sum_{j=1}^{n} g_j - \frac{c}{s} & \text{if } i \in S \\ w - g_i + a\sum_{j=1}^{n} g_j & \text{if } i \notin S \end{cases} . \tag{4.2}$$

Note that by joining the institution, members commit to comply with the decision made collectively in the institution. Members' contribution level is determined by the chosen project, while non-members' is determined outside the institution, individually.

If $S = \emptyset$, i.e. the institution is not implemented, then agent i's payoff is given by

$$\pi_i = w - g_i + a\sum_{j=1}^{n} g_j, \tag{4.3}$$

where all contribution levels are determined individually.

Following the idea of backward induction, let us consider the subgames that start in the contribution stage. Zero contribution is optimal for each agent in the contribution stage if no institution is implemented (this subgame is identical to the public-goods games analyzed at the beginning of this section). If the institution is implemented, members do not make decisions in the final stage, their contribution level is determined in the previous stage by the vote, while it is still optimal for non-members to contribute zero.

A key insight of this model is that, although agents individually have incentives to free-ride in the public-goods game that appear in the final stage, they could increase joint and also individual payoffs by coordinating their contributions with the help of implementing an institution. In equilibrium, members earn $asw - \frac{c}{s}$ if the institution has been implemented and project 1 has been chosen, and $\frac{1}{2}asw - \frac{c}{s}$ if project $\frac{1}{2}$ has been chosen. Due to the incentives to free-ride in the contribution stage, everybody earns w if no institution is implemented. Note that voting for project $\frac{1}{2}$ is a weakly dominated strategy for all members (dominated by voting for 1).

Kosfeld et al. (2009) consider institutions that make their decisions based on unanimity, which implies that each vote is pivotal. The use of the unanimity rule instead of the plurality rule increases the set of equilibria of the game, because the change of a single member's vote will not always affect the collective decision inside the institution when the plurality rule is used (e.g., when all the others vote for the same project). In what follows, we are going to consider equilibria in weakly undominated strategies. This implies that members are going to vote for project 1 as long as the number of members s is such that

$$asw - \frac{c}{s} \geq w. \tag{4.4}$$

Solving for s gives the minimum institution size s^* for members in the implementation stage to vote for project 1 rather than project 0 to dissolve the institution.

Kosfeld et al. (2009) show that for any number of members s there exists a status-quo equilibrium in which no institution is implemented, because of the strong *status-quo* (i.e. project 0) bias of the unanimity rule. Although the plurality rule is less biased towards the status quo, when at all (and at least two) members vote for project 0 a single deviation can not change the outcome. Therefore a similar no-institution equilibrium exists in our game, too.

More importantly, there exist organizational equilibria in which an institution is implemented when $s \geq s^*$. Kosfeld et al. (2009) also show that the institution-formation game has a unique strict subgame-perfect equilibrium in terms of the institution size. Strictness of the equilibrium requires that player play a unique best-response strategy in the equilibrium, i.e. they should be strictly worse off by deviating from the equilibrium strategy. In such an equilibrium, exactly s^* players become members of the institution.[2]

[2] "Strictness of equilibrium yields a clear-cut prediction regarding organisation size: exactly the minimum number of players s^* required for the organisation to be individually profitable form the institution. Unless $s^* = n$, players are thus divided into two proper subsets: those who voluntarily implement the sanctioning institution, hence contributing to the public good, and those who do not participate and do not contribute." (Kosfeld et al., 2009) As long as $s - 1 \geq s^*$, there will exist a player in S who can choose not to join the institution and free-ride instead. Thus in equilibrium only institution of size equal to s^* will be implemented.

In summary, by focusing on subgame-perfect Nash equilibria that are strict in the participation stage and do not include undominated strategies in the implementation stage, we expect that institutions of s^* will be implemented and that members will be required to contribute their entire initial endowment, while non-members will contribute nothing.

4.3 Experimental Procedure

As discussed in there previous section, the core of our experimental study is a ten-player (n=10) linear public-goods game with institution formation adopted from Kosfeld et al. (2009). Five treatments have been designed to explore the effect of different voting systems on contribution levels and institution formation. A sample of the instructions that were used in the experiment is in the Appendix.

In the baseline treatment, players (with initial endowments of 20 points (w=20)) played the public-goods game with voluntary institution formation and used the plurality rule to choose one among the three available public projects inside the institution (PLU3).[3] Participants were informed about the number of people who decided to join the institution in the participation stage. At least two participants were required to form an institution which had a total cost of $c = 2$. In the implementation stage, members of the institution had to decide whether to dissolve the institution and not to force its member to contribute anything to the public good (Project 0), or to force all its members to contribution exactly half of their initial endowment (Project $\frac{1}{2}$: 10 points), or to force all its members to contribute their entire initial endowment to the public good (Project 1: 20 points). Participants did not incur the cost of forming the institution if Project 0 was chosen.

Players who had not joined the institution did not make any decisions at the implementation stage and were not informed about the decision reached by the members of the institution either. They (and also including those who joined the institution if Project 0 was selected) were required to decide how much to contribute to the public goods in the final, contribution

[3] In case of a tie, a project was randomly chosen from those that got the most votes.

stage. The MPCR (a) from contributing to the public good was set to 0.4. Note that, given these parameter values, the minimal institution size for an organisational equilibrium, i.e. for members to vote for implementation (rather than dissolution), is $s^* = 3$.

Table 4.1 offers an overview of all our five treatments containing sequences of the above described game and/or its slightly modified versions.

- In treatment PLU3, participants played the PLU3 game in fixed groups for 20 rounds.
- In treatment PLU3RS, participants first played the PLU3 game in fixed groups for 20 rounds. Then, in newly assigned groups, they played another 20 rounds of the PLU3 game.
- In treatment PLU3SUB, participants first played a subgame of the PLU3 game for 40 rounds (in fixed groups for two sequences of 20 rounds). The subgame that they played was in which all participants were exogenously forced to join the institution at the participation stage of the game. The rest of the game was as in the original PLU3. After round 40, participants were reassigned to groups and played the PLU3 game for another 20 rounds.
- Treatment UNA consisted of two games. Game UNA3 is identical to PLU3 except that the decision of the institution is made by the unanimity voting rule. In game UNA2, the unanimity rule is applied and only two projects (Project 0 and Project 1) are available. Note that our UNA2 game replicates the Kosfeld et al. (2009) design, but the group size is larger (10 people instead of 4). In this treatment, participants played in fixed groups 20 rounds of the UNA3 game, and in reassigned fixed groups another 20 rounds of the UNA2 game.
- The game in treatment MAJ2 differed from PLU3 only in that Project $\frac{1}{2}$ was not available to the members of the institutions. This is why the plurality rule got simplified into a majority rule.

TABLE 4.1: Treatment summary

	PLU3	PLU3RS	PLU3SUB	UNA	MAJ2
# OF SESSIONS	1	1	2	2	1
GAMES	PLU3	PLU3+PLU3	SUB+SUB+PLU3	UNA3+UNA2	MAJ2
# OF PARTICIPANTS	20	20	20	20	20
# OF GROUPS	2	2	2	2	2
# OF ROUNDS	20	20+20	20+20+20	20+20	20
# OF PLU3 OBS.	40	80	40	-	-

20 subjects participated in each session and were randomly assigned into two groups of ten. The experiment was computerized with zTree (Fischbacher, 2007) and participants were not allowed to communicate with each other.

We conducted one session of treatments PLU3, PLU3RS and MAJ2, and two sessions of the treatments PLU3SUB and UNA. All the experiments were run at the experimental laboratory at Waseda University (Tokyo, Japan). In total, 140 subjects participated in our experiment. No one was allowed to participate in more than one session. On average sessions lasted around 90 minutes. Earnings accumulated during the experiments were converted to Japanese yen at the rate of 2 points to 1 yen, and participants earned around 1700 JPY (about $17 including show-up fee). Participants were paid individually and privately at the end of the session.

4.4 Results

In the data analysis, we focus on the PLU3 game, i.e. the public-goods game with voluntary institution formation under the plurality voting rule with three available projects, but report results for all other game, too. The statements and results are based on parametric statistical tests and are significant at least at 5% significance level, unless stated otherwise.

Table 4.2 offers a detailed summary of our experimental results and also shows the main findings from related works in the literature for comparison.

It shows in the first panel of Table 4.2 that, in line with prior evidence, overall initiation rates[4] are close to 100% in all treatments with some fluctuations and variations between initial (first five rounds) and final play (last five rounds).

The implementation rate[5] is presented in the second panel of Table 4.2. From all the initiated institutions on average more than 50% get established in all our treatment with one remarkable exception: in the UNA3 game, the implementation rate is merely 12.5%. Just like in the other treatments, it experiences a significant and important drop from when the first five periods are compared to the last five, but even during the initial round it (30%) is located well below the comparable rates from other treatments. The average institution is composed by 3-4 members in the treatments based on the plurality and majority voting schemes, and by 5+ in the treatments based on the unanimity rule. In the UNA3 game the institution size reached a remarkable maximum of 8 members. (See the third panel of Table 4.2.) We believe that the relatively large number of participants of the institution and the additional available project made reaching a unanimous decision very difficult, especially because communication was not allowed before voting. We shall look into this results with the help of regression analysis later.

The fourth panel of Table 4.2 describes the contribution level. The overall contribution levels to the public good are around 30-40% of the initial endowment. They are constantly very high (roughly 90-100%) inside the institution, and very low (10-30%) with a decreasing trend outside the institution. Efficiency rates[6] across treatments show a very similar picture, given that they are computed with the help of group earnings and the cost

[4]The initiation rate is the proportion of groups (throughout the 20 rounds among all groups) in which at least one participants decides to join the institution in the first stage of the game.

[5]The implementation rate is the proportion of groups (throughout the 20 rounds among all groups that initiated an institution) in which the members of the institution managed to choose (with the help of the imposed voting scheme) a project other than Project 0.

[6]Following Kosfeld et al. (2009), the efficiency rate is defined as $(\sum_i \pi_{observed} - \sum_i \pi_{min})/(\sum_i \pi_{max} - \sum_i \pi_{min})$, where $\sum_i \pi_{observed}$ denotes the observed group earnings, $\sum_i \pi_{min}$ is the theoretical minimum group earnings (200 in all our games), and $\sum_i \pi_{max}$ is the theoretical maximum group earnings (800 in all our games).

of forming an institution is relatively low in our games. See the last panel of Table 4.2.

In sessions PLU3RS and PLU3SUB, participants were required to make various decisions before playing the PLU3 game in the final 20 rounds.[7] (See Table 4.1.) We refer to those results as experienced play even if the experience that participants gained differed. We did not find any significant difference in contribution levels or implementation rates when looking at observations from those games ($p = 0.9411$ for the Wilcoxom-Mann-Whitney test), therefore we pooled them for future analysis and we will refer to them as from experienced play. Similarly, we pooled data from the PLU3 session and the first 20 rounds of session PLU3RS under the name of inexperienced play ($p = 0.3252$ for the Wilcoxom-Mann-Whitney test).

[7] In treatment PLU3RS, they played 20 rounds of the entire PLU3 game, and 2 sets (groups reshuffled after each set) of 20 rounds of a subgame in the treatment PLU3SUB.

TABLE 4.2: Summary of experimental results

	PLU3 POOLED	PLU3 INEXP.	PLU3 EXP.	KOR	UNA2	UNA3	BK	MAJ2	BK
# OF OBSERVATIONS	200	80	120	220	80	80	80	40	80
GROUP SIZE	10	10	10	4	10	10	10	10	10
MPCR	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.6
# OF PROJECTS	3	3	3	2	2	3	2	2	2
VOTING RULE	PLU	PLU	PLU	UNA	UNA	UNA	MAJ	MAJ	MAJ
INITIATION RATE (%)									
AVERAGE	96.00	91.25	99.17	100.00	100.00	100.00	-	90.00	-
AVERAGE (PERIODS 1-5)	98.00	100.00	96.67	100.00	100.00	100.00	-	70.00	-
AVERAGE (PERIODS 16-20)	92.00	80.00	100.00	100.00	100.00	100.00	-	100.00	-
IMPLEMENATION RATE (%)									
AVERAGE	58.85	57.53	59.66	43.18	63.75	12.50	-	52.78	-
AVERAGE (PERIODS 1-5)	71.43	75.00	68.97	-	80.00	30.00	-	42.86	-
AVERAGE (PERIODS 16-20)	45.65	37.50	50.00	-	60.00	5.00	-	50.00	-
INSTITUTION SIZE									
AVERAGE	3.69	3.71	3.67	3.85	5.24	5.00	3.50	3.32	5.06
MAXIMUM	7	6	7	4	8	8	6*	6	7*
AVERAGE (PERIODS 1-5)	3.97	3.67	4.20	-	5.81	5.00	-	3.00	-
MAXIMUM (PERIODS 1-5)	6	6	6	-	8	6	-	4	-
AVERAGE (PERIODS 16-20)	3.38	3.50	3.33	-	4.50	8.00	-	3.00	-

TABLE 4.2: Summary of experimental results (Continued)

	PLU3 POOLED	PLU3 INEXP.	PLU3 EXP.	KOR	UNA2	UNA3	BK	MAJ2	BK
MAXIMUM (PERIODS 16-20)	6	6	5	-	7	8	-	5	-
CONTRIBUTION (% OF INIT. ENDOWMENT)									
AVERAGE	35.66	42.43	31.14	53.00	37.45	26.56	38.40	29.31	57.90
AVERAGE (IN)	94.60	90.38	97.13	-	100.00	100.00	-	100.00	-
AVERAGE (OUT)	22.25	33.27	15.89	-	8.15	23.19	-	15.66	-
AVERAGE (PERIODS 1-5)	47.45	55.63	42.00	-	50.25	43.10	-	32.70	-
AVERAGE (IN, PERIODS 1-5)	94.24	88.18	98.21	-	100.00	100.00	-	32.70	-
AVERAGE (OUT, PERIODS 1-5)	31.95	44.19	22.50	-	8.71	36.70	-	29.31	-
AVERAGE (PERIODS 16-20)	24.53	27.20	22.75	-	30.50	13.38	-	23.35	-
AVERAGE (IN, PERIODS 16-20)	95.07	95.24	95.00	-	100.00	100.00	-	100.00	-
AVERAGE (OUT, PERIODS 16-20)	15.24	22.93	11.83	-	7.50	12.14	-	12.50	-
EFFICIENCY (%)									
AVERAGE	35.47	42.25	30.94	51.00	37.24	25.11	-	29.15	-
AVERAGE (PERIODS 1-5)	47.22	55.38	41.78	-	49.98	43.00	-	32.60	-
AVERAGE (PERIODS 16-20)	24.39	27.10	22.58	-	30.30	13.36	-	23.18	-

NOTE: PLU3 - all PLU3 games with or without experience; PLU3 inexp. - PLU3 games from the first 20 rounds in session PLU3RS; PLU3 exp. - PLU3 games from the last 20 rounds in sessions PLU3RS and PLU3SUB; KOR - results from treatment IF40 in Kosfeld et al. (2009); BK - results from Burger and Kolstad (2009). All variables are as defined in the main text.

4.4.1 The Effect of the Voting Rule

In order to analyze the impact of the voting rule on decisions in the various stages of the game, it is desirable to control for the effect of other variables such as experience, the number of available projects, etc. Table 4.3 reports odds-ratio estimates from logit models on institution-formation and implementation decisions in the first two stages of the game, and coefficient estimates from a linear-regression model on individual contribution decisions in the last stage of the game. Note that all regressions exclude observations from the PLU3SUB game, and in the case of implementation observations are for groups, while in all the other cases they are for participants. This is why the number of observations differs radically in the third numerical column of the table. Individual voting decisions will be analyzed later in the text. All the regressions that analyze individual decisions incorporate a large number of control variables. They were created with the help of the answers participants gave in the post-experimental questionnaire. Most of the questions are related to demographics and attitudes towards cooperation and competition (refer to Appendix for details and the complete list).[8]

As compared to the unanimity rule studied by Kosfeld et al. (2009), the plurality/majority rule serves as a filtering device: only those who *strongly* would like to form an institution join in the first stage given that it is rather difficult to withdraw and abandon the institution later. The unanimity rule completely eliminates the risk involved in the first-stage decision, because each participant has veto power when deciding the future of the institution. Kosfeld et al. (2009) report 100% initiation rate, and that is exactly what we observe in our UNA2 and UNA3 games.

The odds ratios estimated for the PLU./MAJ. dummy, which captures the effect of the voting rule, are not only highly significant, but also important in their size. The use of the plurality/majority rule reduces the odds

[8]Although participants were required to report their major to an open-ended question, we decided to use answer to the more specific questions about whether they studied Economics, Microeconomics, or Game Theory instead. Also, due to problems of multicollinearity we excluded the variable related to membership in associations (and student circles), given that all of the participants reported to be a member of at least one.

of joining to somewhere between a quarter and a half of the odds otherwise.[9] At the same time, it also makes the odds of successfully forming an institution is almost 30 times larger. (See Regressions (1)–(3).) These results are consistent with the column-to-column average comparisons that we described above based on the numbers in Table 4.2. It is still remarkable that the average initiation rate is at least 90% in all our games in spite of the risk of having to follow the majority even if one does not agree to its decision.

As for the end result, i.e. contribution levels and realised efficiency, the use of the plurality/majority voting rule does not constitute a significant change. (See Regressions (4) and (5).) The negative impact of the plurality/majority voting rule on the first-stage joining decisions and its positive impact on successful institution formation practically entirely cancel each other out. The following subsections look further into the details of the observed joining, voting and contributing behavior in order to disentangle the impact of our treatment variables (voting rule, number of projects, and experience).

4.4.2 Institution Formation

The numbers in Table 4.3 indicate a significant, but rather small negative time trend in all decisions, i.e. as time goes by - on average - participants are less likely to join, institutions are less likely to be formed and contributions are likely to decrease (refer to the PERIOD row in Table 4.3). We interpret these effects as the result of getting to know one's interacting partners and their behavior better. That is one kind of experience which is different from the effect of getting to know the rules of the game better. The latter (captured by the EXPERIENCE regressor) does not seem to matter for the joining decision in the first stage, but it does make the implementation of the institution significantly more likely (it induces a threefold increase in the odds). It also reduces the contribution level significantly by roughly 1.75 points. This effect might look surprising at the first sight,

[9] The difference between the two logit regression for joining decision is that the second includes some lagged variables among its regressors.

TABLE 4.3: Regression analysis of strategic behavior (institution formation and contribution)

	(1)	(2)	(3)	(4)	(5)
	JOIN LOGIT, OR	JOIN LOGIT, OR	FORMED LOGIT, OR	CONTR. OLS	CONTR. OLS
PERIOD	0.9833***	0.9731***	0.9576**	-0.1613***	-0.1987***
PLU./MAJ.	0.2580***	0.4944***	28.4223***	-0.0841	0.0649
EXPERIENCE	1.0310	0.8809	3.1547***	-1.7467***	-2.4549***
# OF PROJECTS	1.5168***	1.2071	0.2165***	0.7080***	0.8811***
JOIN (LAG1)		7.3328***			
JOIN (LAG2)		3.9046***			
EARNING (LAG1)		0.9801***			
EARNING (LAG2)		0.9898***			
# JOINED			2.0996***		
INSIDE				15.5316***	5.5710***
PERIOD * INSIDE					0.2409***
EXPERIENCE * INSIDE					3.2091***
INST.SIZE					0.4581***
INST.SIZE^2					-0.0727**
INST.SIZE * INSIDE					1.4490**
INST.SIZE^2 * INSIDE					-0.0519
CONS.	0.2202**	0.8200	0.3889	3.4930**	2.7908*
CONTROLS	YES	YES	NO	YES	YES
R^2	0.1325	0.3920	0.2305	0.6240	0.6370
# of obs.	4000	3600	400	4000	4000

NOTE: coefficient significantly different from zero at ***1%, **5%, *10% significance level. Estimated odds ratios are reported for the logit regressions. Observations from the PLU3SUB are excluded. JOIN - 1 if the participants decides to join the first stage, 0 otherwise; FORMED - 1 if the institution is formed, 0 otherwise; CONTR. - individual contribution level. PLU./MAJ. - 1 if the plurality or majority rule is applied, 0 when unanimity is required; # JOINED - number of participants who join in the first stage; INSIDE - 1 if the participant is a member of the institution, 0 otherwise; INST.SIZE - size of the implemented institution. When indicated, we used answer to the post-experimental questionnaire as CONTROLS.

but we believe it captures the sharp decrease in contribution outside the institution. In other words, it shows that experienced participants (who are more familiar with the game) are more likely to play the dominant strategy outside the institution which indicates not to contribute at all. (Also refer to the fourth panel of Table 4.2.)

The odds of successfully forming an institution are closely (significantly and in an important way) related to the number of participants who show their willingness to join in the first stage of the game. On average, each extra member causes a twofold increase in those odds. For a more accurate picture, we have disaggregated the implementation rates according to institution size in Table 4.4. Recall that in our experimental design the minimum efficient institution size is of three members, and it was technically impossible to form an institution with only one member.

In the PLU3 games, even inefficient institutions are getting implemented of around 50% of the cases. This relatively large rate, which is remarkably close to the case for the minimum efficiency size, might have been inflated by the random tie-breaking rule that would determine the fate of the institution if one member votes for, while the other against its implementation. Among all the initiated institutions with two members (49 in the PLU3 games), only 11 are implemented such that neither of the members votes for Project 0, i.e. to dissolve the institution. That corresponds to an implementation rate of 22.45%.

Even if the implementation rate is far from the ideal 100%, the most popular institution size both for initiated and implemented institutions is three, which is exactly the minimum efficient size. It is when the number of potential members is four or more, that the implementation rate jumps well about 80% and reaches a constant 100% if at least half of the group, i.e. five people, decide to join. The grand coalition of ten was not formed in any of our treatments (or games). The largest institution both initiated and/or implemented has seven members in the PLU3 games, which confirms our conjecture that when the group size is substantially larger than the minimum efficient size, it is unlikely for the grand coalition to form. This is in sharp contrast with the results reported by Kosfeld et al. (2009) for a design where the group size is four and the minimum efficient size is three. They stress that "the majority (on average, around 75 %) of the organizations implemented are grand organizations."

Exogenously determined membership makes the institution formation almost always successful. However, it is worth noticing that there are players (among the 10 participants) who persistently vote to dissolve the institution. The unanimity rule would result in the institution to be dissolved, but the plurality rule does not and that is how participants managed to always achieve full efficiency in those games. Burger and Kolstad (2009), who used a large group size (10 as in our design), a lower MPCR at 0.3 and only two projects, report an average institution size of 3.5, which is somewhat lower than our 3.7 in the PLU3 games. The MAJ2 game has an the average institution size of 3.32, which is lower than the average size of 3.7 in the PLU3 games. We hypothesize that an additional project could have potentially lured more subjects to participate in the first stage, and therefore led to a higher average institution size. This is in line with our regression analysis results in Table 4.3 and the numbers reported in Table 4.4: the unanimity rule makes the initiation easier, but the implementation more difficult, having three (instead of only two) available projects has a similar, but somewhat less pronounced effect.

In the PLU3 game, at an individual level, 42 (21%) out of 200 participants never chose to join in the first stage, and another 70 (35%) participants chose to join five times or less. On the other hand, only 4 (2%) subjects chose to join all the time, and only 18 (9%) participants chose to join 15 times or more. In the games with unanimity rule, i.e. UNA2 and UNA3, 16 (20%) and 12 (15%) out of 80 always stayed out, and another 16 (20%) and 8 (10%) participants chose to join five times or less, respectively. These proportions are not drastically different from the previously discussed ones. However, in the UNA2 and UNA3 games many more participants decided to join almost all the time. 14 (17.5%) and 12 (15%) did so in all 10 rounds, and altogether 32 (40%) and 40 (50%) participants chose to join 15 times or more, respectively.

TABLE 4.4: Relative frequency initiation and implementation, and implementation rate by institution size

INST.		PLU3			PLU3			PLU3			UNA2			UNA3			PLU2	
SIZE	POOLED	POOLED	POOLED	INEXP.	INEXP.	INEXP.	EXP.	EXP.	EXP.									
	INIT.	IMPL.	IMPL.%	INIT.	IMPL.	IMPL.%	INIT.	IMPL.	IMPL.%	INIT.	IMPL.	IMPL.%	INIT.	IMPL.	IMPL.%	INIT.	IMPL.	IMPL.%
0	4.00	-	-	8.75	-	-	0.83	-	-	-	-	-	-	-	-	10.00	-	-
1	13.50	-	-	18.75	-	-	10.00	-	-	-	-	-	-	-	-	12.50	-	-
2	24.50	21.24	48.98	21.25	23.81	58.82	26.67	19.72	43.75	3.75	1.96	33.33	1.25	0.00	0.00	37.50	31.58	40.00
3	28.50	29.20	57.89	25.00	26.19	55.00	30.83	30.99	59.46	10.00	1.96	12.50	3.75	10.00	33.33	20.00	31.58	75.00
4	14.00	22.12	89.29	10.00	19.05	100.00	16.67	23.94	85.00	18.75	11.76	40.00	12.50	30.00	30.00	10.00	15.79	75.00
5	8.50	15.04	100.00	8.75	16.67	100.00	8.33	14.08	100.00	43.75	56.86	82.86	17.50	30.00	21.42	7.50	15.79	100.00
6	6.50	11.50	100.00	7.50	14.29	100.00	5.83	9.86	100.00	13.75	13.73	63.64	37.50	20.00	6.67	2.50	5.26	100.00
7	0.50	0.88	100.00				0.83	1.41	100.00	7.50	9.30	83.33	13.75	0.00	0.00			
8										2.50	3.92	100.00	10.00	10.00	12.50			
9													3.75	0.00	0.00			
10																		
TOTAL	100.00	100.00	58.85	100.00	100.00	57.53	100.00	100.00	59.66	100.00	100.00	63.75	100.00	100.00	12.50	100.00	100.00	52.78

NOTE: PLU3 - all PLU3 games with or without experience; PLU3 INEXP. - PLU3 games from the first 20 rounds in session PLU3RS; PLU3 exp. - PLU3 games from the last 20 rounds in sessions PLU3RS and PLU3SUB. INIT. - initiated institution; IMPL. - implemented institution; IMPL.% - implementation rate.

4.4.3 Voting

Voting, from a game-theoretical point of view, constitutes a rather complicated coordination problem, and especially so in the case of the plurality rule with more than two projects. In all our games, we expect participants to vote for Project 1 if and only if enough, i.e. at least $s^* = 3$, participants join the institution in the first stage. Even if Project $\frac{1}{2}$ in available is some game, it should never be voted according to the previously discussed theoretical argument.

Table 4.5 presents the distribution of votes observed in our experiments according to the size of the initiated institution. The numbers confirm the regression results by showing that, after its initiation, institution formation is especially difficult if unanimity is required and when there are more projects.

It is worth noting that the theoretical threshold for efficient institution formation $s^* = 3$ is barely enough for forming an institution even under the plurality rule. Inexperienced participants needed at least 4 members for at least half of them to vote for Project 1 as Project $\frac{1}{2}$ was particularly popular among them. In the games based on the plurality rule, we observe that as the number of participants who join in the first stage increases, members feel more confident (or optimistic) about the prospects of an institution and the vast majority of them vote for Project 1. Project $\frac{1}{2}$ is only popular among inexperienced participants and when the initiated institution is small. It also could be the case that participants do realize that the institution could be efficiently formed, but are reluctant to do so, when too many participants have decided to stay out and it is very likely that they are going to free ride on the contributions forcefully extracted by the institution. Project $\frac{1}{2}$ might then represent a compromise solution to the members' dilemma.

Interestingly, Project $\frac{1}{2}$ turns out to be popular - and practically irrespectively from the institution size so - whenever it is available under unanimity rule. In that case, theoretical efficiency is not enough as a coordination device when more than two projects are to be considered by voters.

TABLE 4.5: Distribution of votes (%) by institution size

INST. SIZE	PLU3 POOLED			PLU3 INEXP.			PLU3 EXP.			UNA2			UNA3			PLU2		
	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1
2	51	14	35	47	26	26	53	8	39	67	-	33	0	50	50	63	-	37
3	42	16	43	38	18	43	43	14	42	33	-	67	22	11	67	33	-	67
4	14	19	67	6	44	50	18	9	74	18	-	82	13	30	58	19	-	81
5	11	11	79	6	17	77	14	6	80	3	-	97	10	30	60	13	-	87
6	6	10	83	8	19	72	5	2	93	6	-	94	7	37	56	0	-	100
7	0	0	100				0	0	100	2	-	98	3	29	69			
8										0	-	100	2	23	75			
9													0	11	89			
10																		

NOTE: PLU3 - all PLU3 games with or without experience; PLU3 inexp. - PLU3 games from the first 20 rounds in session PLU3RS; PLU3 exp. - PLU3 games from the last 20 rounds in sessions PLU3RS and PLU3SUB.

A surprising result is that even in the PLU3SUB game, where all participants are forced to join the institution, not all of them vote for Project 1, which is the *obvious* optimal strategy. Although 77.75% of the votes went for Project 1 in those (sub)games, 19.75% went for Project 0 and 2.5% to Project $\frac{1}{2}$.

4.4.4 Contributions and Efficiency

Efficiency in our games is tightly linked to individual contribution levels and successful (efficient) institution formation. We define efficiency as the proportion of the additional profit above the absolute minimum level that a group of participants achieve from the additional profit that they could have achieved.

The last two panels in Table 4.2 reveal that the average contribution level is below half of the initial endowment, somewhere around 27-43%, and therefore observed efficiency is also rather low (25-42%). For a deeper analysis and comparison across treatments we turn to the regression results in the last two columns of Table 4.3.

Just like in the *usual* linear public-goods experiment (e.g. Fehr and Gächter 2000), our regression reveals a significant, although not very steep, negative trend in contribution levels. Experience with a similar game or subgame reduces contributions more radically, by 1.7 points (out of 20 points of initial endowment) on average. This effect is similar to what appears as a restart effect in other studies (Andreoni, 1988), given that experience here means having played the same or a very similar game with a different group of participants. Even if contribution levels decay with time, a restart usually sends them back to a relatively high level which is typically below to the one at the previous start. Another interpretation, based on the coefficient estimates in the regression (5) of Table 4.3, is that experience makes game-theoretical incentives clearer. Therefore participants outside the institution contribute less with experience (it is a dominant strategy for them to free-ride after all), while experienced participants inside the institution contribute more with experience (as they realise that voting for Project 1 is the best strategy whenever the institution is large enough).

Interestingly, the voting rule has no significant effect on contributions. Its impacts on initiation and institution formation, with opposing signs as discussed before, entirely cancel each other out. It is important to underline though that contribution levels are significantly and remarkably higher inside the institution. If we look at Regression (4) in table 4.3 which does not control for the size of the implemented institution, and add the constant and the coefficient that measures the average difference between inside and outside the institution, we get 19 points. That is nearly all of the 20 points that were available for each participant in each period to decide about. However, comparing the average contribution level when no institution is formed, the failure of institution formation has a much bigger negative effect on the contribution level in game UNA2 than in game PLU3 ($p < 0.0001$ for the Wilcoxom-Mann-Whitney test).

As for the effect of institution size, it is quite different inside and outside the institution (as shown by the four related coefficient estimates). Inside the institution, it is positive and significant, and it is likely to be explained by the previously discussed voting and institution-formation behavior which converges toward the equilibrium Project 1 as the number of members in the institutions is getting larger. Outside the institution it has a much smaller impact which is positive for institutions with 6 or less members and negative for larger ones. In other words, these results show that relatively small institutions do generate a significant positive (although moderate in size) externality on contribution levels outside their boundaries. We do not have a plausible explanation for it, but the number of projects also seem to have a significant but rather small positive effect on contribution levels.

4.5 Conclusions

We experimentally studied a linear public-goods game with endogenous institution formation. Participants were asked to decide whether to join an institution before deciding on their contribution to the public good. Members of the institution were required to contribute the amount the institution members voted for. While previous studies have shown that institution formation has a positive effect on the overall efficiency (Kosfeld et al., 2009, Burger and Kolstad, 2009), our experimental design focuses on the effect of the voting rule used by the institution for collective decision-making. In the literature and numerous real-life situations, the unanimity rule is the most commonly used voting rule. However, following the breakdown of various international negotiations (environmental summits, security and peace talks) and the increasing dispute in the European Union about the sluggishness of its decision-making process, unanimity rule has been criticised for being too restrictive and therefore hindering cooperation. In our experimental design, we replaced the unanimity rule by a more flexible majority/plurality rule and investigated its effects on institution initiation, institution formation and contribution levels.

Firstly, our observations show that the majority/plurality rule significantly decreases the initiation rate, but at the same time also significantly increases the implementation rate of institution. These two effects cancel each other out in the end, and suggest that the choice of the voting rule (unanimity or majority/plurality) does not matter for the average contribution level or efficiency.

Secondly, our experimental design with an increased group size (from the 4 in the literature to 10) shows that the grand coalition, even if it would be an efficient outcome, does not form. This result puts previous experimental evidence (especial the one presented by Kosfeld et al., 2009) into different light and suggests that what matters for institution formation is whether the to-be-formed institution has enough potential members or not as compared to the theoretically minimum efficient institution size. Whether that is closer or farther from the size of the grand coalition is of secondary importance. Interestingly, that theoretical minimum size is not enough of a guarantee for an institution to be formed in the experimental laboratory.

Chapter 5

A Non-cooperative Approach to the Talmud Solution for the Bankruptcy Games[1]

5.1 Introduction

In a bankruptcy problem, every creditor has a certain claim over a perfectly divisible estate which is insufficient to grant all the claims. The bankruptcy problem captures the essence of most rationing situations, such as the execution of a will to disburse the insufficient property to the beneficiaries, the distribution of the liquidation of a firm among its creditors, the collection of a certain amount of tax among tax payers with different responsibilities, etc.. Such problems, arising from the Talmud, are first studied by O'Neil (1982) and Aumann and Maschler (1985). In order to allocate the estate "fairly" among all the players, several rules have been proposed, including the proportional rule (PRO), constrained equal awards rule (CEA), constrained equal losses rule (CEL), etc.. Aumann and Maschler (1985) first formulated a solution from the Talmud (which we call the Talmud solution (TAL)) which, restricting the creditors by their half claims, applies CEA to the problem first and the excess (if any) is divided using CEL. Most studies

[1]This chapter is based on a joint work with Yuan Ju.

since then take an axiomatic approach. (See Thomson (2003) and Moulin (2002) for comprehensive reviews.)

In this chapter, we take a non-cooperative approach to the bankruptcy problems. The aim is to provide a strategic game that yields the Talmud solution in subgame perfect equilibrium (SPE). We first propose two games, each of which has a unique SPE outcome that coincides with the CEA and CEL allocation respectively. Then the two games are combined such that the unique SPE outcome is the TAL allocation.

Studying the bankruptcy problems as non-cooperative games provides additional support for the rules alongside their axiomatic considerations. As early in O'Neil (1982)'s study, a claim game was proposed where any Nash equilibrium gives the minimum overlapping rule allocation. Atlamaz et al.(2011) extended O'Neill (1982)'s claim game and variations of their game result in proportional division in equilibrium.

In this study, we mainly focus on the non-cooperative approach to the CEA, CEL and TAL. Chun (1989) defined a game where players propose order preservative rules and the limit point of a process of adjustment coincides with the CEA allocation. Herrero (2001) studied the dual game of the above mechanism that reaches the CEL allocation as the unique Nash equilibrium outcome. Sonn (1992) studied a game of demand (in a modified alternating offer bargaining game style), where the SPE outcome converge to the CEA allocation when the discount factor goes to 1. Different from the above approach, the game proposed in this study reaches the CEA (CEL) in SPE rather than the limit point of the procedure. It allows the players to achieve the desirable allocation in finite steps.

In the line of non-cooperative method to TAL, existing mechanisms rely on the contested-garment(CG) consistency. Serrano (1995) proposed a bargaining game whose SPE yields the TAL. Dagan et al.(1997) extended the mechanism to a class of consistent rules. In these two mechanisms, a disagreement point corresponding to the CG solution of two-person reduced game is imposed to guarantee the desired outcome. This is a major difference from the strategic games described in this chapter where the

TAL allocation as the SPE outcome does not rely on the specification of consistency.

The games proposed in this chapter are inspired by the famous cut-and-choose mechanism in fair division (Brams and Taylor (1996)). When two people share a (homogeneous) cake, they try to make the division as fair as possible. One way to solve the dispute is to randomly choose one of the players to cut the cake, while the other player has the right to choose first which piece he gets. Thus, the best strategy for the player who cut the cake is to cut it exactly in haves. The unique SPE outcome is an equal division of the cake. This mechanism is extended to a n-player game preserving its fairness consideration, while applied to a bankruptcy game by restricting that no player get a payoff higher than his claim. As conventionally in the literature of bankruptcy problems, the estate and claims are assumed as common knowledge.

We first introduce a game where the player with the highest claim is appointed as the executor and makes a proposal which is an *efficient* allocation of the estate. Following the proposal, players, from the one with the lowest claim, sequentially choose from the elements of the vector as their payoffs.[2] However, if at any point, a player's choice is higher than his claim, he only gets his claim and the executor makes a new proposal with respect to the remaining estate for the players who have not received their payoffs. The game ends when the executor takes the remaining estate after all other players have received their payoff. It can be shown that the unique SPE outcome of this game is the CEA allocation of the corresponding bankruptcy problem. A similar game, where the deficit between the estate and total claim is distributed by the executor, yields the CEL allocation as the unique outcome in all SPE. A combination of the two games gives the TAL allocation as the unique SPE outcome. In each of the above games, the outcome could be reached in finite steps and the desired allocation is the outcome in all SPE.

[2]Players with lower claims are given priority to choose early on in the game, while the player who chooses last is given the power to make the division. The privilege of dividing the estate and choosing early is spread among the players.

The rest of the chapter is organised as follows. The next section introduces the problem and the allocation rules. In section 3, we present the games and the main results. The last section concludes with several extensions of the current model.

5.2 Bankruptcy Problems and the Talmud Solution

Let $N = \{1, 2, \ldots, n\}$ be the finite set of players. For each $i \in N$, let $c_i \in \mathbb{R}_+$ denote player i's claim and $c = (c_i)_{i \in N}$ the vector of claims. $E \in \mathbb{R}$ is the perfectly divisible estate to be divided among all players. A bankruptcy problem is a pair (E, c), such that $0 < E < \sum c_i$.[3] Without loss of generality, we assume players are ordered according to their claims, that is, for players $1, 2, \ldots, n$, we have $0 \leq c_1 \leq c_2 \cdots \leq c_n$. The order is randomly decided among players with equal claims. An allocation in a bankruptcy problem is an n-tuple $x(E, c) = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, with $\sum x_i = E$ and $0 \leq x_i \leq c_i$. An allocation rule is a function that assigns a unique allocation to each bankruptcy problem.

Constrained equal awards rule (CEA) divides the estate equally among all players, subject to the constraint that no player gets more than his claim.

$$CEA_i(E, c) = \min\{\alpha, c_i\}, \text{with } \alpha \text{ solves} \sum \min\{\alpha, c_i\} = E.$$

Constrained equal losses rule (CEL) assigns equal loss to each player, subject to the constraint that no player receives less than zero.

$$CEL_i(E, c) = \max\{0, c_i - \beta\}, \text{with } \beta \text{ solves} \sum \max\{0, c_i - \beta\} = E.$$

TAL connects CEA and CEL, but half of each player's claim is used as the bound of award or loss. CEA is applied to the problem first, and if there is excess of the estate after the distribution, CEL is applied.

[3]For simplicity of notation, $\sum$ stands for $\sum_{i=1}^{n}$ unless specified otherwise.

$$TAL_i(E,c) = \begin{cases} CEA_i(E, \frac{c_i}{2}) & \text{when } \sum c_i \geq 2E; \\ c_i - CEA_i(\sum c_i - E, \frac{c_i}{2}) & \text{when } \sum c_i < 2E \ . \end{cases}$$

The above equation could be written as the following:

$$TAL_i(E,c) = \begin{cases} CEA_i(E, \frac{c_i}{2}) & \text{when } \sum c_i \geq 2E; \\ \frac{c_i}{2} + CEL_i(E - \sum \frac{c_i}{2}, \frac{c_i}{2}) & \text{when } \sum c_i < 2E. \end{cases}$$

The equivalency between the above two expressions could be seen by realising when $\sum c_i < 2E$,

$$\begin{aligned} TAL_i(E,c) &= c_i - CEA_i(\sum c_i - E, \frac{c_i}{2}) \\ &= c_i - \min\{\beta, \frac{c_i}{2}\}, \text{s.t.} \sum \min\{\beta, \frac{c_i}{2}\} = \sum c_i - E \\ &= c_i + \max\{-\beta, -\frac{c_i}{2}\}, \text{s.t.} \sum \max\{-\beta, -\frac{c_i}{2}\} = -(\sum c_i - E) \\ &= \frac{c_i}{2} + \max\{\frac{c_i}{2} - \beta, 0\}, \\ &\quad \text{s.t.} \sum \max\{\frac{c_i}{2} - \beta, 0\} = \sum \frac{c_i}{2} - \sum c_i + E = E - \frac{1}{2} \sum c_i \\ &= \frac{c_i}{2} + CEL_i(E - \sum \frac{c_i}{2}, \frac{c_i}{2}) \end{aligned}$$

5.3 The Mechanism

We propose a multi-step extensive form (divide-and-choose) game whose unique SPE outcome coincides with the TAL allocation.

Consider a bankruptcy problem (E,c). We first define the game in each step.

5.3.1 Game Γ^1

Game $\Gamma^1(E,c)$ is a divide-and-choose game with respect to the bankruptcy problem (E,c). Denote the payoff for player i in game Γ^1 as π_i^1.

The game consists of $n+1$ stages: the opening stage 0 where the executor makes a proposal of how the estate would be divided, $n-1$ subsequent stages where in each stage $s \in \{1,2,\ldots,n-1\}$, each agent s sequentially makes choice of his share of the estate from the *opening* proposal, and finally a stage n where the executor gets his payoff.

For each stage s, there are two proposals: one before player s makes his choice which we call the "opening proposal" and denote as A_o^s (the subscript o stands for "opening") and the one at the end of stage s which we call the "closing proposal" and denote as A_c^s (the subscript c stands for "closing"). There is only a closing proposal in stage 0.

Stage 0 Player n, who is the player with the highest claim, is appointed as the executor.[4]

Player n proposes an allocation $A_c^0 = (x_1^0, x_2^0, \ldots, x_n^0) \in \mathbb{R}^n$. The game proceeds to stage 1.

For $s = 1,2,\ldots,n-1$, the stages proceed in the following way.

Stage s Player s chooses an element from stage s' opening proposal $A_o^s = (x_s^s, x_{s+1}^s, \ldots, x_n^s)$, where $A_o^s = A_c^{s-1}$. From stage 1, players sequentially make choices from the opening proposal. The opening proposal of each stage is the same as the closing proposal of the previous stage. Define player s' choice as $\theta_s = x_i^s, i = s, s+1, 2, \ldots, n$. The smaller of his choice or his claim is his payoff.

- If $\theta_s \leq c_s$, player s leaves the game with payoff $\pi_s^1 = \theta_s$. Stage s' closing proposal is $A_c^s = (y_{s+1}^s, y_{s+2}^s, \ldots, y_n^s) \in \mathbb{R}^{n-s}$, where A_c^s is a

[4]If there are more than one of players have the highest claim, an executor is randomly chosen among such players. Without loss of generality, player n is always chosen as the executor. Since it is always the executor who makes the proposal, we use the word "executor" and "proposer" interchangeably.

projection of A_o^s to $\mathbb{R}^{n-s}$, such that

$$|\{j|y_j^s = \theta_s, \text{for some } j = s+1, 2, \ldots, n\}| \\ = |\{k|x_k^s = \theta_s, \text{for some } k = s, \ldots, n\}| - 1. \quad (5.1)$$

If his choice is smaller or equal to his claim, the closing proposal of stage s is the same as its opening proposal except that player s' choice is removed from it, which means the number of such elements that equal to θ_s is one less in the closing proposal, while all the other elements are the same.

The game proceeds to the next stage.

- If $\theta_s > c_s$, player s leaves the game with payoff $\pi_s^1 = c_s$. Player n makes a closing proposal $A_c^s = (y_{s+1}^s, y_{s+2}^s, \cdots, y_n^s) \in \mathbb{R}^{n-s}$ with respect to the remaining players $s+1, s+2, \ldots, n$ and the remaining estate $E - \sum_{j=1}^{s} \pi_j^1$ such that

$$\sum_{i=s+1}^{n} y_i^s = E - \sum_{j=1}^{s} \pi_j^1.$$

The game proceeds to the next stage.

Stage n Player n leaves the game with payoff

$$\pi_n^1 = \max\{0, \min\{c_n, E - \sum_{i=1}^{n-1} \pi_i^1\}\}.$$

We show that game Γ^1 has a unique SPE outcome that coincides with the allocation assigned by CEA.

Define the estate remaining at stage 1 as $E_1 = E$, at stage $s = 2, \ldots, n-1$ for the $n-s+1$ players as $E_s = E - \sum_{i=1}^{s-1} \pi_i^1$.

The following lemma shows that at any stage, if the remaining estate is not enough to award all the remaining players the amount as much as the lowest claim among the remaining players, the opening proposal of the stage must be an equal division of the remaining estate among the remaining players. By reverse induction on the number of remaining players, we show that under the condition specified above, if the opening proposal is not equal division, the proposer would end up with a lower payoff. Thus, any proposer's strategy that leads to a non-equal division at such stage is not part of the SPE.

Lemma 5.1. If $E_s < (n-s+1)c_s$ at any stage $s, s = 1, 2, \ldots, n-1$, in all SPE, $A_o^s = (x_s^s, x_{s+1}^s, \ldots, x_n^s) \in \mathbb{R}^{n-s+1}$, where $x_i^s = \frac{E_s}{n-s+1}, i = s, \ldots, n$.

Proof. The proof is done by reverse induction on s.

We first show that the lemma holds for $s = n-1$. When $s = n-1$, the lemma says if $E_{n-1} < 2c_{n-1}$, then in all SPE, $A_o^{n-1} = (x_{n-1}^{n-1}, x_n^{n-1})$, where $x_{n-1}^{n-1} = x_n^{n-1} = \frac{E_{n-1}}{2}$.

Given the proposal A_o^{n-1}, either choice of player $n-1$ is a best response. The game ends with player $n-1$ and player n get payoffs $\pi_{n-1}^1 = \pi_n^1 = \frac{E_{n-1}}{2}$.

Now suppose the proposal is $(A_o^{n-1})' = ((x_{n-1}^{n-1})', (x_n^{n-1})')$, where $(x_{n-1}^{n-1})' = \frac{E_{n-1}}{2} - \varepsilon, (x_n^{n-1})' = \frac{E_{n-1}}{2} + \varepsilon, \varepsilon \in (0, \frac{E_{n-1}}{2}]$. Player $n-1$ is at best response to choose $(x_n^{n-1})'$. $(\pi_{n-1}^1)' = \min\{\frac{E_{n-1}}{2} + \varepsilon, c_{n-1}\} > \frac{E_{n-1}}{2}$. The game ends with player n gets payoff $(\pi_n^1)' = E_{n-1} - (\pi_{n-1}^1)' < \frac{E_{n-1}}{2} = \pi_n^1$. At any stage prior to stage $n-1$, player n would have at least one opportunity to make a proposal to guarantee A_o^{n-1} at equilibrium, giving him a higher payoff then the current situation. Hence $(A_o^{n-1})'$ cannot be part of SPE.

Assume the lemma holds for $s = 2$. We show that the lemma holds for $s = 1$.

When $s = 1$, the lemma says if $E_1 < nc_1$, then in all SPE, $A_o^1 = (x_1^1, x_2^1, \ldots, x_n^1) \in \mathbb{R}^n$ where $x_i^1 = \frac{E_1}{n}, i = 1, 2, \ldots, n$. Given the proposal A_o^1, for any $i \in N$, any choice of player i is a best response. No new proposal is made by player n in the following stages. The only outcome following proposal A_o^1 is $\pi_i^1 = \frac{E_1}{n}$.

Suppose $(A_o^1)' = ((x_1^1)', (x_2^1)', \ldots, (x_n^1)'), (x_1^1)' \leq (x_2^1)' \leq \cdots \leq (x_n^1)'$ (at least one inequality holds) and $\sum (x_i^1)' = E_1$, we must have $(x_1^1)' < \frac{E_1}{n}$ and $(x_n^1)' > \frac{E_1}{n}$. Player 1's payoff at his best response is $(\pi_1^1)' = \min\{(x_n^1)', c_1\} > \frac{E_1}{n}$.

Then at stage $s = 2$, $(E_2)' = E_1 - \min\{(x_n^1)', c_1\} < \frac{(n-1)E_1}{n} < (n-1)c_1 \leq (n-1)c_2$. By the assumption that the lemma holds at $s = 2$, player n's equilibrium payoff is $(\pi_n^1)'$ when $(A_o^2)' = (\frac{(E_2)'}{n-1}, \ldots, \frac{(E_2)'}{n-1})$, $(\pi_n^1)' = \frac{(E_2)'}{n-1} < \frac{E_1}{n} = \pi_n^1$.[5] Player n can increase his payoff by deviating from $(A_o^1)'$ and making a proposal $A_c^0 = A_o^1$. Therefore, proposal $(A_o^1)'$ cannot constitute an SPE. □

The next theorem shows that game Γ^1 has a unique SPE outcome which coincides with the allocation assigned by CEA in corresponding bankruptcy problem. We show that in all SPE, the players whose claims are lower than the average of the remaining estate over remaining players would have their claims as their payoffs given any proposal. Until all the remaining players have higher claims than the average of the remaining estate, then by lemma 5.1, they would be have the average remaining estate as their payoff.

Theorem 5.2. For any bankruptcy problem (E, c), the unique SPE outcome of game $\Gamma^1(E, c)$ is $\pi_i^1 = CEA_i(E, c)$.

Proof. The proof is done by induction on the number of players.

We first show that the theorem holds for $|N| = 2$.

Case 1, $E < 2c_1$. $CEA_i(E, c) = \frac{E}{2}, i = 1, 2$. By lemma 5.1, the only SPE proposal at stage 1 is $A_o^1 = (x_1^1, x_2^1)$, where $x_1^1 = x_2^1 = \frac{E}{2}$. Any choice of player 1 constitutes an SPE. All SPE yield the same outcome $\pi_i^1 = \frac{E}{2} = CEA_i(E, c), i = 1, 2$.

Case 2, $E \geq 2c_1$. $CEA(E, c) = (c_1, E - c_1)$. Any proposal $A_o^1 = (x_1^1, x_2^1)$, such that $x_1^1 + x_2^1 = E$ at stage 1 constitutes an SPE in this case. In any SPE, Player 1's choice is $\theta_1 = \max\{x_1^1, x_2^1\} \geq c_1$ and his payoff is $\pi_1^1 = \min\{\theta_1, c_1\} = c_1$. In all SPE, the payoffs are $\pi_1^1 = c_1$ and $\pi_2^1 = E - c_1$, which are the CEA allocation.

Assume the theorem holds for $|N| = n - 1$. We show the theorem also holds for $|N| = n$.

[5] $(\pi_n^1)' - \pi_n^1 = \frac{E_2}{n-1} - \frac{E_1}{n} = \frac{nE_2 - (n-1)E_1}{n(n-1)} = \frac{E_1 - n \cdot \min\{x_n', c_1\}}{n(n-1)}$. Since $E_1 < nx_n, E_1 < nc_1, E_1 - n \cdot \min\{x_n', c_1\} < 0$, which means $(\pi_n^1)' < \pi_n^1$.

Case 1, if $E \leq nc_1, CEA_i(E,c) = \frac{E}{n}$ for all $i = 1,2,\ldots,n$. By lemma 5.1, the only SPE proposal starting from stage 1 is $A_o^1 = (x_i^1)_{i=1,\ldots,n}$ where $x_i = \frac{E}{n}$. All SPE yield the same outcome $\pi_i^1 = \frac{E}{n} = CEA_i(E,c)$.

Case 2, if $nc_1 < E < \sum c_i, CEA_i(E,c) = \min\{c_i, \alpha\}$, where $\sum \min\{c_i, \alpha\} = E$.

Consider the following strategy profile $\mathscr{G}$. At stage 0, player n propose $A_c^0 = (x_i^0)_{i=1,2,\ldots,n}$, where $x_i^0 = c_i$ for $i = 1,\ldots,q, q < n$ and $x_i^0 = \alpha$ for $i = q+1,2,\ldots,n$, with $\sum x_i^0 = E$ and q and α solve $c_q \leq \alpha < c_{q+1}$. At every stage s, player s' choice is $\theta_s = x_s^0 \in A_o^s$. Players' payoffs are $\pi_i^1 = c_i$ for $i = 1,\ldots,q$ and $\pi_i^1 = \alpha = \frac{E-\sum_{i=1}^q c_i}{n-q+1}$ for $i = q+1,2,\ldots,n$, which coincides with the CEA allocation.

To see that the above strategy constitute an SPE, note that every player's choosing strategy is his best response. For player $i = 1,\ldots,q$, deviating from the current θ_i would give no more than his current payoff. Player n's proposal at stage 0 is also his best response. Any other proposal would lead to a payoff of c_i for player $i, i = 1,\ldots,q$, leaving $E_{q+1} = E - \sum_{i=1}^q c_i < (n-q+1)c_{q+1}$. By lemma 5.1, player n's payoff is maximised with$A_o^{q+1} = (x_i^{q+1})_{i=q+1,2,\ldots,n}, x_i^{q+1} = \frac{E_{q+1}}{n-q+1}$ at $\pi_n^1 = \frac{E_{q+1}}{n-q+1}$, which is no higher than the payoff he can get with proposal A_o^1. Thus, no player has incentive to deviate from the strategy profile $\mathscr{G}$ which is an SPE, whose outcome coincides with the CEA allocation.

It could also be shown that all SPE yield the same outcome which is the CEA allocation. In all SPE, because $E > nc_1$ and the proposal A_o^1 must be efficient as in $\sum x_i^1 = E$, there must exist an x_i^1 such that $x_i^1 \geq c_1$. Player 1's best response at stage 1 is $\theta_1 = x_i^1 \geq c_1$ for any $i = 1,2,\ldots,n$. His payoff is $\pi_1^1 = c_1$. At stage 2, $E_2 = E - c_1$. By the assumption that theorem holds for $n-1$ players, all SPE starting from stage 2 give the same payoff for player $2,\ldots,n$ which is $\pi_i^1 = \min\{c_i, \alpha\}, i = 2,\ldots,n, \sum_{i=2}^n \pi_i^1 = E - c_1$. Because $\min\{c_2, \alpha\} > c_1$[6] the SPE outcome for n players could be expressed as $\pi_i^1 = \min\{c_i, \alpha\} = CEA_i(E,c), i = 1,2,\ldots,n, \sum \pi_i^1 = E$.

□

[6] If $\alpha < c_2$, then we must have $\alpha = \frac{E-c_1}{n-1} > c_1$; if $\alpha \geq c_2$, by assumption, $c_2 > c_1$.

5.3.2 Game Γ^2

Game $\Gamma^2(E,c)$ is a divide-and-choose game with respect to the bankruptcy problem (E,c). In order for the estate to be distributed, each creditor has to lose part of his claim. The procedure is similar to that of Game $\Gamma^1(E,c)$. Game $\Gamma^2(E,c)$ distribute the total loss, i.e. the difference between the estate and the sum of claims, among all creditors so that every creditor's payoff is his claim minus his distributed loss.

Define the total loss for all creditors N as $L_1 = \sum c_i - E$. Denote the payoff for player i in game Γ^2 as π_i^2.

The procedure resembles that in game Γ^1. The game consists of $n+1$ stages: an opening stage 0 and n subsequent stages, each involving one player.

Stage 0 Player n, who is the player with the highest claim, is appointed as the proposer.

Player n makes a proposal $B_c^0 = (x_1^0, x_2^0, \dots, x_n^0) \in \mathbb{R}^n$ such that $\sum x_i = L_1$. The game proceeds to stage 1.

For $s = 1, 2, \dots, n-1$, the stages proceed in the following way.

Stage s Player s chooses an element from stage s' opening proposal $B_o^s = (x_s^s, x_{s+1}^s, \dots, x_n^s)$, where $B_o^s = B_c^{s-1}$. Define player s' choice as $\xi_s = x_i^s, i = s, s+1, 2, \dots, n$.

- If $\xi_s \leq c_s$, player s leaves the game with payoff $\pi_s^2 = c_s - \xi_s$. Stage s' closing proposal is $B_c^s = (y_{s+1}^s, y_{s+2}^s, \dots, y_n^s) \in \mathbb{R}^{n-s}$, where B_c^s is a projection of B_o^s to $\mathbb{R}^{n-s}$, such that

$$|\{j | y_j^s = \xi_s, \text{for some } j = s+1, 2, \dots, n\}| \\ = |\{k | x_k^s = \xi_s, \text{for some } k = s, \dots, n\}| - 1. \quad (5.2)$$

  The game proceeds to the next stage.

- If $\xi_s > c_s$, player s leaves the game with payoff $\pi_s^2 = 0$. Player n makes a proposal $B_c^s = (y_{s+1}^s, y_{s+2}^s, \dots, y_n^s) \in \mathbb{R}^{n-s}$, such that $\sum_{i=s+1}^n y_i^s = L_1 - \sum_{j=1}^s (c_j - \pi_j^2)$. The game proceeds to the next stage.

Stage n Player n leaves the game with payoff

$$\pi_n^2 = \max\{0, \min\{c_n, E - \sum_{i=1}^{n-1} \pi_i^2\}\}.$$

The following lemma and theorem show that the only SPE outcome of game $\Gamma^2(E,c)$ coincides with the allocation assigned by CEL.

Theorem 5.3. For any bankruptcy problem (E,c), the unique SPE outcome of game $\Gamma^2(E,c)$ is $\pi_i^2 = CEL_i(E,c)$.

The proofs are done similarly to those of Lemma 5.1 and Theorem 5.2. We first need two Lemmas to facilitate the proof of Theorem 5.3.

Lemma 5.4. Given two bankruptcy problem (E,c) and (E',c) and the CEL allocation for each problem as $CEL_i(E,c) = \min\{0, c_i - \beta\}, \text{s.t.} \sum \min\{0, c_i - \beta\} = E$ and $CEL_i(E',c) = \min\{0, c_i - \beta'\}, \text{s.t.} \sum \min\{0, c_i - \beta'\} = E'$ for players $i = 1, 2, \ldots, n$. If $E' < E$, then $\beta' > \beta$.

Proof. The claim vector is $c = (c_1, c_2, \ldots, c_n), c_1 \leq c_2 \leq \cdots \leq c_n$. Without loss of generality, assume $c_j \leq \beta < c_{j+1}$ and $c_k \leq \beta' < c_{k+1}$ with $j,k = 1, 2, \ldots, n-1$. Explicitly:

$$\overbrace{0 + \cdots + 0}^{j} + c_{j+1} - \beta + c_{j+2} - \beta + \cdots + c_n - \beta = E$$

$$\overbrace{0 + \cdots + 0}^{k} + c_{k+1} - \beta' + c_{k+2} - \beta' + \cdots + c_n - \beta' = E'$$

First, with $E' < E$, we must have $j \leq k$.

If $j > k$, and because $j,k = 1, 2, \ldots, n-1$, it must be $j \geq k+1$. Then we have $c_j \geq c_k \geq c_{k+1}$. $c_k \leq \beta' < c_{k+1} \leq c_j \leq \beta < c_{j+1}$, which means $\beta' < \beta$ when $j > k$.

$$\begin{aligned} E' - E &= (c_{k+1} - \beta' + c_{k+2} - \beta' + \cdots + c_n - \beta') \\ &\quad - (c_{j+1} - \beta + c_{j+2} - \beta + \cdots + c_n - \beta) \\ &= (c_{k+1} - \beta' + c_{k+2} - \beta' + \cdots + c_j - \beta') + (n-j)(\beta - \beta') \end{aligned}$$

The first part of the above equation is positive and $j < n$. $E' < E$ requires $E' - E < 0$. So we must have $\beta - \beta' < 0, i.e. \beta < \beta'$, which is a contradiction of $\beta' < \beta$ when $j > k$.

If $j = k$,

$$\begin{aligned} E' - E &= (c_{k+1} - \beta' + c_{k+2} - \beta' + \cdots + c_n - \beta') \\ &\quad - (c_{j+1} - \beta + c_{j+2} - \beta + \cdots + c_n - \beta) \\ &= (n-j)(\beta - \beta') \end{aligned}$$

$E' - E < 0$, so $\beta' > \beta$.

If $j < k$, and because $j,k = 1,2,\ldots,n-1$, it must be $j+1 \leq k$. Then we have $c_j \leq c_{j+1} \leq c_k \leq c_{k+1}$. $c_j \leq \beta < c_{j+1} \leq c_k \leq \beta' < c_{k+1}$, which means $\beta < \beta'$ when $j < k$.

$$\begin{aligned} E' - E &= (c_{k+1} - \beta' + c_{k+2} - \beta' + \cdots + c_n - \beta') \\ &\quad - (c_{j+1} - \beta + c_{j+2} - \beta + \cdots + c_n - \beta) \\ &= -(c_{j+1} - \beta + c_{j+2} - \beta + \cdots + c_k - \beta) + (n-k)(\beta - \beta') \end{aligned}$$

$\beta < \beta'$ satisfies $E' - E < 0$. □

Define the loss at stage $s = 2,\ldots,n-1$ for the $n-s+1$ players as $L_s = L_1 - \sum_{i=1}^{s-1}(c_i - \pi_i^2)$.

Lemma 5.5. If $L_s < (n-s+1)c_s$ at any stage $s, s = 1,2,\ldots,n-1$, in all SPE, $B_o^s = (x_s^s, x_{s+1}^s, \ldots, x_n^s) \in \mathbb{R}^{n-s+1}$, where $x_i^s = \frac{L_s}{n-s+1}, i = s,\ldots,n$.

Proof. If $L_s \leq (n-s+1)c_s$, given the proposal $B_o^s = (x_i^s)_{i=s,\ldots,n}$, with $x_i^s = \frac{L_s}{n-s+1}$, any choice the remaining players is a best response. Since $\frac{L_s}{n-s+1} \leq c_s \leq c_{i,i>s}$, the only outcome following this proposal is $\pi_i^2 = c_i - \frac{L_s}{n-s+1}, i = s,\ldots,n$.

Now suppose the proposal at stage s is $(B_o^s)' = (x_s^s, x_{s+1}^s, \ldots, x_n^s), x_s^s \leq x_{s+1}^s \leq \cdots \leq x_n^s$ (at least one inequality holds) and $\sum x_i^s = L_s$, we must have $x_1^s < \frac{L_s}{n-s+1} < c_s$ and $x_n^s > \frac{L_s}{n-s+1}$. Player s is at best response to choose x_s^s. His payoff is $\pi_s^2 = c_s - x_s^s$.

If $\nexists i \in \{s+1, 2, \ldots, n\}$, s.t.$x_i^s > c_i$, then player n has no chance to make new proposal until he gets payoff $(\pi_n^2)' = c_n - x_n^s < c_n - \frac{L_s}{n-s+1} = \pi_n^2$. Player n gets a lower payoff following the proposal $(B_o^s)'$. At any stage prior to stage s, player n would have at least one opportunity to make a proposal to guarantee the proposal B_o^s at stage s, giving him a higher payoff then the current situation. Hence $(B_o^s)'$ cannot be part of SPE.

If $\exists i \in \{s+1, 2, \ldots, n\}$, s.t.$x_i^s > c_i$, suppose

- For $s+1 < i < n-1, x_i^s \leq c_i$ and for $i = n, x_i^s > c_n$. At equilibrium, player n gets payoff of 0, which is lower than π_n^2. Thus, $(B_s^o)'$ does not constitute an SPE.

- For $s+1 < i < n-2, x_i^s \leq c_i$ and for $i = n-1, x_i^s > c_{n-1}$. Since $x_n^s \geq x_{n-1}^s$, we have $x_n^s > c_{n-1}$. $L_{n-1} = x_{n-1}^s + x_n^s$. At equilibrium, player $n-1$ gets payoff of 0. $L_n = x^s 1 + x_2^s - c_{n-1} > x_n^s > \frac{L_s}{n-s+1}$. Player n gets payoff $(\pi_n^2)' = c_n - \min\{L_n, c_n\} < \pi_n^2$. Thus, B_s' does not constitute an SPE.

- For $s+1 < i < n-3, x_i^s \leq c_i$ and for $i = n-2, x_{n-2}^s > c_{n-2}$. Since $x_n^s \geq x_{n-1}^s \geq x_{n-2}^s$, we have $x_n^s \geq x_{n-1}^s > c_{n-2}$. $L_{n-2} = x_{n-2} + x_{n-1} + x_n$. At equilibrium, player $n-2$ gets payoff of 0. $L_{n-1} = L_{n-2} - c_{n-2} > x_{n-2}^s + x_n^s$. The next proposal from player n satisfies $\min\{y_{n-1}, y_n\} \leq \frac{L_{n-1}}{2}$. Player $n-1$ is at best response to choose $\min\{y_{n-1}, y_n\}$. $L_n = L_{n-1} - \min\{y_{n-1}, y_n, c_{n-1}\} > \frac{L_{n-1}}{2} > \frac{x_{n-1}^s + x_n^s}{2} > c_{n-2} > c_s > \frac{L_s}{n-s+1}$. Player n gets payoff $(\pi_n^2)' = c_n - \min\{L_n, c_n\} < \pi_n^2$. Thus, $(B_o^s)'$ does not constitute an SPE.

- By similar argument, we can show that for any $s+1 < i \leq n$, if $\exists i \leq n$, s.t. $x_{s+1}^s > c_{s+1}$, $(B_o^s)'$ does not constitute an SPE.

□

Now we are ready to prove Theorem 5.3.

Proof. The proof is done by induction on the number of players.

We first show that the theorem holds for $|N| = 2$.

Case 1, $c_2 - c_1 < E$, i.e. $L_1 = c_1 + c_2 - E < 2c_1$. $CEL_i(E,c) = c_i - \frac{L_1}{2}$. By lemma 5.5, the only SPE proposal starting from $s = 1$ is $B_o^1 = (x_i^1)_{i=1,2}, x_i^1 = \frac{L_1}{2}$. Either choice of player 1 constitute an SPE. The only SPE outcome is $\pi_i^2 = c_i - \frac{L_1}{2} = CEL_i(E,c)$.

Case 2, $E \le c_2 - c_1$, i.e. $L_1 \ge 2c_1$. $CEL_i(E,c) = (0,E)$. Any proposal $B_1^o = (x_i^1)_{i=1,2}$ such that $x_i^1 \ge c_1$ constitutes an SPE in this case. For player n, any other proposal $(B_1^o)'$ would lead to smaller loss for player 1, given player 1's best response is always to choose the smaller share of loss, which would lead to a lower payoff for player 2. So $(B_1^o)'$ cannot be par to SPE. The payoff in all SPE for player 1 is $\pi_1^2 = 0$ and that for player 2 is $\pi_2^2 = E$.

Assume that the theorem holds for $|N| = n-1$. We show that it also holds for $|N| = n$.

Case 1, if $\sum c_i - nc_1 < E$, i.e. $L_1 = \sum c_i - E < nc_1, CEL_i(E,c) = c_i - \frac{L_1}{n}$ for all $i = 1,2,\ldots,n$. By lemma 5.5, the only SPE proposal starting from stage 1 is $B_1^o = (x_i)_{i=1,2,\ldots,n}, x_i = \frac{L_1}{n}$. Any strategy of player $1,2,\ldots,n$ constitute an SPE. All SPE yield the same outcome $\pi_i^1 = c_i - \frac{L_1}{n} = CEL_i(E,c)$.

Case 2, if $\sum c_i - nc_1 \ge E$, i.e. $L_1 = \sum c_i - E \ge nc_1, CEL_i(E,c) = c_i - \min\{c_i, \beta\}$, where $\sum \min\{c_i, \beta\} = L_1$.

Consider the following strategy $\mathscr{G}$. At stage 0, player n propose $B_c^0 = (x_i^0)_{i=1,2,\ldots,n}$, where $x_i^0 = c_i$ for $i = 1,\ldots,q, q < n$ and $x_i^0 = \beta$ for $i = q+1,2,\ldots,n$, with $\sum x_i^0 = E$ and q and β solve $c_q \le \beta < c_{q+1}$. At every stage s, player s's choice is $\xi_s = x_s^0 \in B_o^s$. Players' payoffs are $\pi_i^2 = 0$ for $i = 1,\ldots,q$ and $\pi_i^2 = c_i - \beta = c_i - \frac{\sum_{i=q+1}^n c_i - E}{n-q+1}$ for $i = q+1,2,\ldots,n$, which coincides with the CEL allocation.

To see that the above strategy constitute an SPE, note that for player $i = 1,\ldots,q$, choosing any other share would give no more than 0. At stage $q+1$, $L_{q+1} = L_1 - \sum_{i=1}^q c_i = \sum_{i=q+1}^n \beta < (n-q)c_{q+1}$. By lemma 5.5, in all SPE, $B_o^{q+1} = (x_i^{q+1})_{i=q+1,2,\ldots,n}, x_i^{q+1} = \frac{E_{q+1}}{n-q+1}$ the outcome is $\pi_i^2 = c_i - \frac{L_{q+1}}{n-q+1}$ for player $j = q+1,2,\ldots,n$.

Thus, no player has incentive to deviate from the strategy profile $\mathscr{G}$ which is an SPE, whose outcome coincides with the CEL allocation. Next, we show that all SPE yields the CEL allocation.

Because $L_1 > nc_1$, in all SPE, $B_c^0 = (x_i^0)_{i=1,2,\ldots,n}, x_i^0 \geq c_1$. Player 1 is at best response to choose x_i^0 for any i at stage 1 and gets payoff is 0. Then at stage 2, $L_2 = L_1 - c_1$, the total estate left is still E and the total claim becomes $\sum_{i=2}^{n} c_i$. By the assumption that theorem holds for $n-1$ players, all SPE starting from $s = 2$ give the same payoff for player $i = 2, \ldots, n$ which is $\pi_i^2 = c_i - \min\{c_i, \beta\}, \sum_{i=2}^{n} \pi_i^2 = E$. Because $\min\{c_2, \beta\} > c_1$, the SPE outcome for n players could be expressed as $\pi_i^2 = c_i - \min\{c_i, \beta\}, i = 1, 2, \ldots, n, \sum \pi_i^2 = E$. Specifically, $\pi_i^2 = \min\{0, c_i - \beta\}$ and since $E > 0, \pi_n^2 > 0$.

Suppose in a proposal $(B_c^0)'$, there exist a $x_j^0 < c_1$, at stage 1, player 1 has best response $\xi_1 = x_j^0$ and gets payoff $(\pi_1^2)' = c_1 - x_j^0$. Then at stage 2, $(L_2)' = L_1 - x_j^0$, the total estate left is $E' = E - c_1 + x_j^0$ and the total claim is $\sum_{i=2}^{n} c_i$. By the assumption that theorem holds for $n-1$ players, the SPE outcome starting from stage 2 gives the payoff $\pi_i^2 = c_i - \min\{c_i, \beta'\}, i = 2, \ldots, n, \sum_{i=2}^{n} \pi_i^2 = E' < E$. In this subgame, player n gets a payoff no greater than $\pi_n^2 = \min\{0, c_n - \beta'\}$. Since $E' < E$, then we must have $\beta' > \beta$.(By lemma 5.4.) So $(\pi_n^1)' < \pi_n^1$. Player n has an incentive to deviate from $(B_c^0)'$ to B_c^0. Any proposal other than B_c^0 cannot be part of the SPE. The only SPE outcome is $\pi_i^2 = \min\{0, c_i - \beta\}$ such that $\sum \pi_i^2 = E$.

□

5.4 The Non-Cooperative Game Γ

Game $\Gamma(E, c)$ is a combination of game Γ^1 and game Γ^2. The game has four steps. First, the estate is double. Then players distribute awards by playing the Game $\Gamma^1(E, c)$. If there is any excess left, the players distribute losses by playing the Game $\Gamma^2(E, c)$. In the end, all players get half of their total payoff of these two games to balance the budget of the original estate. Consider any bankruptcy problem (E, c).

Step 0 The total estate is doubled. Let $\hat{E} = 2E$.

Step 1 Game $\Gamma^1(\hat{E},c)$ is played with respect to the bankruptcy problem $(\hat{E},c)$. The allocation for player i in step 1 is his payoff π_i^1 in game $\Gamma^1(\hat{E},c)$.[7] The game continues to Step 2.

Step 2 Game $\Gamma^2(\hat{E}-\sum\pi_i^1,c)$ is played with respect to the bankruptcy problem $(\hat{E}-\sum\pi_i^1,c)$. The allocation for player i in step 2 is his payoff π_i^2 in game $\Gamma^2(\hat{E}-\sum\pi_i^1,c)$.[8]

Step 3 Player i takes the final payoff $\pi_i = \frac{\pi_i^1+\pi_i^2}{2}$.

The next theorem shows that Game $\Gamma(E,c)$ has a unique SPE outcome which coincides with the Talmud solution of the bankruptcy problem (E,c). First, it is shown that if the estate is not enough to grant every player half of her claim, the doubled estate will be fully distributed by the end of step 1. Second, if the estate is more than half of the total estate, in step 1, every player's payoff is his claim. Finally, the payoff for every player is the allocation assigned by the Talmud solution.

Theorem 5.6. For any bankruptcy problem (E,c), the unique SPE outcome of game $\Gamma(E,c)$ is the allocation assigned by Talmud rule for the problem (E,c).

5.4.1 Proof of Theorem 5.6

Before the proof of Theorem 5.6, we first need to show the homogeneity of the Talmud solution.

Lemma 5.7 (Homogeneity of CEA and CEL).

$$CEA(\lambda E,\lambda c) = \lambda CEA(E,c)$$

$$CEL(\lambda E,\lambda c) = \lambda CEL(E,c)$$

Proof.

$$CEA_i(E,c) = \min\{c_i,\alpha\}, \text{with } \alpha \text{ solves} \sum \min\{c_i,\alpha\} = E$$

[7] The restriction that Γ^1 is played with respect to a strict bankruptcy problem is relaxed. $\hat{E} < \sum c_i$ is not required.

[8] The restriction that Γ^2 is played with respect to a positive estate is relaxed. $\pi_i^2 = 0$ if $\hat{E}-\sum\pi_i^1 = 0$.

There is a unique α satisfying the above equation.

$$
\begin{aligned}
CEA_i(\lambda E, \lambda c) &= \min\{\lambda c_i, \alpha'\}, \text{s.t.} \sum \min\{\lambda c_i, \alpha'\} = \lambda E \\
&= \lambda \min\{c_i, \frac{\alpha'}{\lambda}\}, \\
&\qquad \text{s.t.} \sum \lambda \min\{c_i, \frac{\alpha'}{\lambda}\} = 2E, i.e. \sum \min\{c_i, \frac{\alpha'}{\lambda}\} = E
\end{aligned}
$$

By the uniqueness of α, we must have $\alpha = \frac{\alpha'}{\lambda}$.

We have

$$
\begin{aligned}
CEA(\lambda E, \lambda c) &= \lambda \min\{c_i, \alpha\}, \text{s.t.} \sum \min\{c_i, \alpha\} = E \\
&= \lambda CEA(E, c)
\end{aligned}
$$

The proof of $CEL(\lambda E, \lambda c) = \lambda CEL(E, c)$ could be done similarly. □

The following result immediately follows.

$$
TAL_i(E, c) = \begin{cases} \frac{1}{2} CEA_i(2E, c) & \text{when } \sum c_i \geq 2E; \\ \frac{1}{2}(c_i + CEL_i(2E - \sum c_i, c)) & \text{when } \sum c_i < 2E. \end{cases}
$$

Now we are ready to prove Theorem 5.6.

Proof. The proof is done by a series of claims.

Claim 1: if $E \leq \frac{\sum c_i}{2}$, in all SPE, $\hat{E} - \sum \pi_i^1 = 0$ in Step 2. Suppose there are two SPE strategy profiles $\mathscr{G}$ and $\mathscr{G}'$. By following $\mathscr{G}$, $\hat{E} - \sum \pi_i^1 = 0$ in Step 2 and the payoff for any player i is $\pi_i = \frac{\pi_i^1}{2}$, while by following $\mathscr{G}'$, $\hat{E} - \sum \pi_i^1 > 0$ and player's payoff is $(\pi_i)' = \frac{(\pi_i^1)' + (\pi_i^2)'}{2}$.

In $\mathscr{G}'$, at the last stage of Step 1, it must be the case that $(\pi_n^1)' = c_n$.[9] So $\pi_n^1 \leq (\pi_n^1)'$. It implies that $\sum_{i=1}^{n-1} \pi_i^1 > \sum_{i=1}^{n-1} (\pi_i^1)'$.[10]

There must be at least one player $j, j \neq n$ whose allocation in step 1 is lower in $\mathscr{G}'$ than in $\mathscr{G}$, i.e. $c_j \geq \pi_j^1 > (\pi_j^1)'$. Suppose player j is the only such player. At step 1, game Γ^2 is played with respect to $((\pi_j^1 - (\pi_j^1)') - (c_n - \pi_n^1), c)$. Player j's payoff is $(\pi^2 j)' < (\pi_j^1 - (\pi_j^1)') - (c_n - \pi_n^1) < \pi_j^1 - (\pi_j^1)'$. His final payoff is $(\pi_j)' = \frac{(\pi_j^1)' + (\pi_j^2)'}{2} < \frac{\pi_j^1}{2} = \pi_j$. So player j has an incentive to deviate from $\mathscr{G}'$. $\mathscr{G}'$ cannot be an SPE.

Claim 2: if $E \geq \frac{\sum c_i}{2}$, in all SPE, the allocation in step 1 for every player is $\pi_i^1 = c_i$. Suppose there are two SPE strategy profiles $\mathscr{G}$ and $\mathscr{G}'$. By following $\mathscr{G}$, $\pi_i^1 = c_i$ for any i, while by following $\mathscr{G}'$, there exist at least one player j such that $(\pi_j^1)' < c_j$. Because the proposal at any stage in step 1 must be efficient, we have $j \neq n$. Without loss of generality, assume player j is the only such player. In $\mathscr{G}$, the game in step 2 is game Γ^2 with respect to $(\hat{E} - \sum c_i, c)$. Player j's allocation in step 2 is $\pi_j^2 = \max\{c_j - \beta, 0\}$, where β solves $\sum \max\{c_i - \beta, 0\} = \hat{E} - \sum c_i$. Player j's final payoff is $\pi_j = \frac{c_j + \pi_j^2}{2}$. In $\mathscr{G}'$, the game in step 2 is game Γ^2 with respect to $(\hat{E} - \sum c_i + (c_j - (\pi_j^1)'), c)$. His allocation in step 2 is $(\pi_j^2)' = \max\{c_j - \beta', 0\}$, where β' solves $\sum \max\{c_i - \beta', 0\} = \hat{E} - \sum c_i + (c_j - (\pi_j^1)')$. His final payoff is $\pi_j' = \frac{(\pi_j^1)' + (\pi_j^2)'}{2}$. It can be easily shown that $(\pi_j^2)' - \pi_j^2 < c_j - (\pi_j^1)'$.[11] So player j must have an incentive to deviate from $\mathscr{G}'$ which is not an SPE.

Claim 3, in all SPE, the final payoff for every player in game Γ coincides with his payoff assigned by Talmud rule in bankruptcy problem (E, c). By claim 1 and theorem 5.2, for all bankruptcy game (E, c) such that $E \leq \frac{\sum c_i}{2}$, the final payoff for any player i is $\pi_i = \frac{\pi_i^1}{2} = \frac{CEA_i(2E, c)}{2} = TAL_i(E, c)$. The last equality follows from the homogeneity of TAL.

By claim 2, for all bankruptcy game (E, c) such that $E > \frac{\sum c_i}{2}$, the allocation in step is $\pi_i^1 = c_i$. By theorem 5.3, in step 2 the allocation is $\pi_i^2 = CEL_i(2E - \sum c_i, c)$. The final payoff for any player i is $\pi_i = \frac{\pi_i^1 + \pi_i^2}{2} =$

[9] $\hat{E} - \sum \pi_i^1 > 0 \Leftrightarrow \hat{E} - \sum_{i=1}^{n-1} (\pi_i^1)' - (\pi_n^1)' > 0 \Leftrightarrow \hat{E} - \sum_{i=1}^{n-1} (\pi_i^1)' > (\pi_n^1)' > 0$. Because $(\pi_n^1)' = \max\{0, \min\{c_n, E - \sum_{i=1}^{n-1} (\pi_i^1)'\}\}$, we must have $(\pi_n^1)' = c_n$.

[10] At $\mathscr{G}$, if $\pi_n^1 = c_n$, we must have $\hat{E} - \sum_{i=1}^{n-1} \pi_i^1 = c_n$. Since $\hat{E} - \sum_{i=1}^{n-1} (\pi_i^1)' > c_n$, which means $\hat{E} - \sum_{i=1}^{n-1} (\pi_i^1)' > \hat{E} - \sum_{i=1}^{n-1} \pi_i^1$.

[11] $(\pi_j^2)' - \pi_j^2 < c_j - (\pi_j^1)'$.

$\frac{c_i+CEL_i(2E-\sum c_i,c)}{2} = TAL_i(E,c)$. The last equality follows from the homogeneity of TAL.

The idea is that the estate is doubled at the beginning of the game to support players working with their claims (rather than half of their claims). The final payoff is halved in order to balance the budget.

□

5.4.2 The Uniqueness of SPE

In the above model, there exist multiple SPE which lead to the same outcome. A simple modification of the rules could achieve the unique SPE.

First assume that a player always prefers the share equal to his claim in the proposal rather than any share higher, since choosing any higher share would also give the same amount as their claims. In any stage $s > 0$, if the proposer makes a new proposal, he bears a cost of μ. μ is a pure loss which will not be awarded to any other players. Explicitly:

In game Γ^1_μ, at each stage $s > 0$, there is a potential cost μ^1_s for player n.

- If $\theta_s \leq c_s$, $\mu^1_s = 0$.
- If $\theta_s > c_s$, $\mu^1_s = \varepsilon, \varepsilon > 0$.

Then at stage n, player n leaves the game with payoff

$$\pi^1_n = \max\{0, \min\{c_n, E - \sum_{i=1}^{n-1} \pi^1_i\} - \sum_{i=1}^{n-1} \mu^1_i\}.$$

In game Γ^2_μ, at each stage $s > 0$, there is a potential cost μ^2_s for player n.

- If $\xi_s \leq c_s$, $\mu^2_s = 0$.
- If $\xi_s > c_s$, $\mu^2_s = \varepsilon, \varepsilon > 0$.

Then at stage n, player n leaves the game with payoff

$$\pi_n^2 = \max\{0, \min\{c_n, E - \sum_{i=1}^{n-1} \pi_i^2\} - \sum_{i=1}^{n-1} \mu_i^2\}.$$

The game Γ_μ is a combination of game Γ_μ^1 and Γ_μ^2.

It imposes a cost on the proposer for redistribution during the game. In order to minimise the cost, the proposer would make the proposal as the desired allocation at stage 0 in each step. The game Γ_μ has a unique SPE whose outcome is the Talmud solution of the corresponding bankruptcy problem.

5.5 Concluding Remarks

In this chapter, we introduce a non-cooperative game whose unique SPE outcome is the Talmud solution of the corresponding bankruptcy problem. Not only it provides a profounding foundation for the Talmud solution, but also due to the close connection of the Talmud solution and the prenucleolous in TU game, it could potentially be a step towards the noncooperative approach of the solution in TU game.

The current model requires player n to make a redistribution every time a player makes choice higher than his claim. The timing of the redistribution can be modified without changing the outcome. A game Γ^d is defined similarly as Γ only that the differences between every player's choice and his claim are accumulated until the end of the game. After player n receives his payoff, the accumulated estate/losses is shared among all players whose claims have not been satisfied/sacrificed completely, following the rules of game Γ^2. It can be shown that game Γ^d yields the TAL solution as the only SPE outcome.

Chun et al. (2001) proposed a reversed Talmud rule which applies CEL first and if there is excess after the initial distribution, CEA is applied. By switching the order of the two games in step 2 and 3, our model yields a unique SPE outcome that coincides with the reverse Talmud rule.

Furthermore, the model is not restricted to bankruptcy problem. For surplus sharing problems, where the estate to be distributed is bigger than the total of claims, the model could be modified accordingly and reach solutions with distinctive features. The mechanism for the surplus sharing problem (E,c), where $E > \sum c_i$ is a multi-step game. Starting from step 0, let stage $k+1$ be the final stage.

If in every step game Γ^1 is played,then the SPE outcome for the game is $\pi_i = kc_i + CEA_i(E - k\sum c_i, c)$. According to this solution, the estate is shared proportional to the players' claims up to a point where the remaining estate is not enough to share another round. Then the remaining is distributed by CEA. If in every stage game Γ^2 is played,then the SPE outcome for the game is $\pi_i = kc_i + CEL_i(E - k\sum c_i, c)$. Similarly, the problem is reduced to a bankruptcy problem by awarding all players k times of their claims until the remaining estate is less than the sum of all the claims. Then the remaining is distributed by CEL.

Chapter 6

Conclusion

As put by Fehr and Fischbacher (2004) "It is not possible to understand the peculiarities and the forces behind human cooperation unless we understand social norms." The thesis consists experimental and theoretical studies on social norms and cooperation.

The four chapters include three experimental studies respectively on the Golden Rule in the context of the ultimatum game, on trust and expected-norm in the context of the trust game and on the cooperative behaviour in the context of the public goods game; and a theoretical study that builds a link between the cooperative and non-cooperative interpretation of the Talmud solution in the bankruptcy problems.

The first two chapters focus on effects of social norms on people's behaviour in strategic situations. Chapter 2 discusses the Golden Rule in a simple bargaining situation. I experimentally test whether people do not treat others in ways that they themselves find not agreeable, which is called role-reversal consistency. The results show that over three quarters of the subjects are role-reversal consistent. Such subjects are more accurate predictors of whether others accept the way they treat them and also earn more money than role-reversal inconsistent subjects. Furthermore, regression analysis shows that role-reversal consistency is not the result of strategic behaviour, in the sense of maximisation of one's expected monetary earnings given one's beliefs about one's opponent's behaviour. I also find a higher rate of role-reversal consistency under the response method than under the strategy method. In Chapter 3, I explore the phenomenon of people's expectation of social norms and their own behaviour in the context of trust and trustworthiness and investigate how people's trust and trustworthiness are connected to their beliefs about their opponents. The results confirm that about half of the subjects carry out their actions based on their beliefs of what the others would do in their positions. Moreover, subjects' behaviour as trustees are significantly influenced by their experience and feedback of playing the trustors role. The experience of playing the trustees role, even without feedback information, helps the trustors make a better prediction of the trustees' behaviour.

In next two chapters, I study the way people cooperate and how to sustain the cooperative results for all the people involved, using both theoretical and experimental methods.

Chapter 4 experimentally explores the mechanisms that make people more willing to cooperate and increase the overall welfare in a public goods game. I show that compared to the unanimity rule, the majority/plurality rule significantly decreases the institution initiation rate, but at the same time also significantly increases the implementation rate of institutions. In the end, as the two effects cancel each other out, the choice of the voting rule does not significantly affect the average contribution level or efficiency.

In Chapter 5, I theoretically study a well-established cooperative solution for the bankruptcy problem. I first propose two games that yield the constraint equal awards allocation and constraint equal losses allocation respectively as their unique sub-game perfect equilibrium (SPE) outcomes. I then integrate the two games to achieve the Talmud solution in SPE. Finally, modifications to the games are introduced such that the game's SPE is unique and the procedure could be extended to a family of "Talmud-related" solutions in bankruptcy problems and surplus sharing problems.

Besides the results presented above, two design features in experiments from the first two chapters also contribute to the methodology of experimental study of games.

First of all, the three-group-three-part design used in Chapter 2 and 3 structurally guarantees a role-reversal in a two-role game. It makes sure subjects only interact with the same opponent once. The three-group-three-part design could be a useful device when a strict role-reversal matching is required and reputation and reciprocity needs to be suppressed in the game play. It also allows for the study of the effect of experience and feedback information on decision making and beliefs. The experience is specific to that of playing in the opposite role rather than the role that a subject is currently playing.

Secondly, in the belief elicitation task used in Chapter 2 and 3, subjects state probability distribution over certain events instead of making a guess of which event is likely to happen. It allows subjects to state equal probability over two or more events which might represent their view of the situation more accurately. The elicitation of subjects' probability distribution is commonly used to assess subject's risk attitude. However, in studies involving social norms, it is rarely used due to its complex nature. But in the afore mentioned two experiments, we have shown that the stated probability distribution allows us to have a closer look of how subjects evaluate the situation.

Appendix A

Instructions and Screenshots for Experiment in Chapter 2

A.1 Instructions – Introductory remarks (common to both treatments)

WELCOME!

PLEASE WAIT UNTIL THE EXPERIMENTER TELLS YOU TO START

You are about to participate in an experiment in decision-making. Universities have provided the funds for this experiment.

In this experiment we will first ask you to read instructions that explain the decision scenarios you will be faced with. Next, you will be asked to make decisions that will allow you to earn money.

Your monetary earnings will be determined by your decisions and the decisions of other participants in the experiment. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session. Your earnings will be kept strictly anonymous.

It is important that you remain silent and do not look at other people's work. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, exclaim out loud, etc., you will be asked to leave and you will forfeit your earnings. Thank you.

The experiment consists of three parts, parts I, II and III. You will participate in two parts. In each part in which you participate, you will anonymously interact with other participants in the room. For the entire experiment today, you will not interact with any participant more than once. Thus, if you interact with a participant in one part, you will not interact with him/her in the other part in which you participate. The decisions that you make in a part will NEITHER influence the decisions you will be faced with NOR the participants you will interact with in the next part in which you participate.

We will now randomly determine the two parts in which you will participate.

You will participate in parts $<?>$ and $<?>$.

A.2 Instructions – Feedback Treatment

Instructions for Proposers - Offer stage

In this part, you will be asked to decide how to split £7 with each of ten other participants. That is, for each pairing with another participant, you will propose £X for you, and £(7-X) for her/him (X has to be a whole number between 0 and 7, i.e., (0, 1, . . . , 6, 7)).

For each pairing, upon being informed of how you want to split the £7, the other participant will either accept or not accept your proposed split. If the other participant accepts your proposed split you will receive £X, and s/he will receive £(7-X). If the other participant does not accept your proposed split, you will receive £0, and the other participant will receive £0.

In summary, you will make one proposal that will be sent to ten other participants. Each of them will separately and independently decide whether to accept or reject your proposal.

After your proposal has been made and accepted or rejected by each of the ten participants who receive it, you will be informed about each of their decisions.

Two of the ten outcomes will be chosen randomly at the end of the session, and you will be paid the sum of your earnings in them. Thus, the chance that you will be paid for a particular outcome is one in five. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

Are there any questions? Please do not talk with others during the experiment.

Instructions for Responders

In this part, you will receive ten proposals, each from a different participant, on how to split £7 with her/him. That is, for each of ten pairings with other participants, you will receive a proposal of £Y for you, and £(7-Y) for her/him (Y has to be a whole number between 0 and 7, i.e., (0, 1, . . . , 6, 7)).

For each pairing, upon being informed of how the other participant wants to split the £7, you will either accept her/his proposed split, or will not accept it. If you accept the proposed split you will receive £Y, and s/he will receive £(7-Y). If you do not accept the proposed split, you will receive £0, and the other participant will receive £0.

In summary, you will receive ten proposals made separately and independently by ten different participants. You will then decide whether to accept or reject each of the ten proposals you receive.

After you have made a decision on each of the ten proposals you receive, each of the other participants will be informed of the outcome of his or her own proposal (but not that of others').

Two of the ten outcomes will be chosen randomly at the end of the session, and you will be paid the sum of your earnings in them. Thus, the chance that you will be paid for a particular outcome is one in five. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

Are there any questions? Please do not talk with others during the experiment.

Instructions for Idle-Subjects

You are not participating in this part.

Please be patient and wait until this part is over.

A.3 Instructions – No-Feedback Treatment

Instructions for Proposers - Offer stage

In this part, you will be asked to decide how to split £7 with each of ten other participants. That is, for each pairing with another participant, you will propose £X for you, and £(7-X) for her/him (X has to be a whole number between 0 and 7, i.e., (0, 1, . . . , 6, 7)).

For each pairing, the other participant will not be shown your proposed split, but instead will be asked to accept or not accept each of the eight feasible splits ((£0 for you, and £7 for her/him), (£1 for you, and £6 for her/him), . . . , (£7 for you, and £0 for her/him)). upon being informed of how you want to split the £7, the other participant will either accept or not accept your proposed split. If the other participant accepts your proposed split you will receive £X, and s/he will receive £(7-X). If the other participant does not accept your proposed split, you will receive £0, and the other participant will receive £0.

In summary, you will make one proposal that will be sent to ten other participants. Each of them will separately and independently have to decide whether they would accept or reject each of the feasible proposed splits.

After your proposal has been made, and each of the ten other participants to whom you are paired has decided whether they would accept or reject each of the feasible proposed splits, your proposal will be matched to each of their accept/reject decisions for your proposed split, and an outcome for each pairing will be determined. You will only be informed about the outcomes at the end of the session.

Moreover, at the end of the session, two of the ten outcomes will be chosen randomly, and you will be paid the sum of your earnings in them. Thus, the chance that you will be paid for a particular outcome is one in five. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

Are there any questions? Please do not talk with others during the experiment.

Instructions for Responders

In this part, each of ten participants will make a proposal on how to split £7 between you and her/him. That is, for each of ten pairings with other participants, each of the 10 participants paired to you will make a proposal of £Y for you, and £(7-Y) for her/him (Y has to be a whole number between 0 and 7, i.e., (0, 1, . . . , 6, 7)).

Since you will not be informed how each of the other participant wants to split the £7 between you and her/him, you will be asked to decide whether you would accept or not accept each of the eight feasible splits ((£0 for you, and £7 for her/him), (£1 for you, and £6 for her/him), . . . , (£7 for you, and £0 for her/him)). For each of the feasible splits if you accept it you will receive £Y, and s/he will receive £(7-Y), if £Y is the amount proposed to you. If you do not accept the proposed split, you will receive £0, and the other participant will receive £0.

In summary, ten different participants will separately and independently each make a proposal on how to split £7 with you. Since you will not see the proposals you will be asked to decide for each of them whether to accept or reject each of the feasible splits.

After you have decided whether you would accept or reject each of the feasible splits, each of the ten proposals will be matched to your accept/reject decision for the proposed split, and an outcome for each pairing will be determined. You will only be informed about the outcomes at the end of the session.

Moreover, at the end of the session, two of the ten outcomes will be chosen randomly, and you will be paid the sum of your earnings in them. Thus, the chance that you will be paid for a particular outcome is one in five. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

Are there any questions? Please do not talk with others during the experiment.

Instructions for Idle-Subjects

You are not participating in this part.

Please be patient and wait until this part is over.

A.4 Instructions for Proposers' Belief Elicitation stage (common to both treatments)

To finish this part we now ask you to give us your ESTIMATE about the chance a randomly chosen participant would accept each of the different feasible splits. Specifically, we ask you for each feasible split, how likely do you think that some other participant would accept it?

You will be asked to answer with a percentage number. If you are absolutely sure that that other participant would ACCEPT the proposed split, then you would want to answer with 100%. If you think it is absolutely certain that the other participant would REJECT the proposed split, then you would want to answer with 0%. If you are less certain that that other participant would ACCEPT the proposed split, then you would want to respond with an intermediate percentage number, reflecting what you think. A higher number would indicate a stronger tendency towards acceptance, and a lower number would indicate a stronger tendency towards rejection.

You will be rewarded for the accuracy of your estimates, as follows:

First, we will randomly select two proposed splits (drawn from the pool of proposals made by all the participants making proposals, including you) received by two participants who were asked to reject or accept them.

Second, for each of the two proposed splits selected, we compute your reward which is equal to 100 points (each point is worth one pence), minus a number "L" (short for "loss") that indicates how well your estimate indicates the decision made by the participant who faced that proposed split.

This number L is determined in several simple steps. The first step is to identify the decision of the participant who received the proposal, i.e., we look up whether s/he accepted or rejected that proposed split. If it was ACCEPT, we take the difference between 100 and your estimate (which we call "ESTIMATE" in the following formula).

Then, this difference is multiplied by itself, and then multiplied by 0.01, yielding the number L.

If, on the other hand, the actual value of that other participant was REJECT, then we simply take your ESTIMATE, multiplied by itself and then by 0.01, to arrive at the number L. Expressed as a formula, your earnings from the estimate are therefore given by:

$100 - L$, where:

- if the other participant's decision was ACCEPT: $L = 0.01 \times (100 - \text{ESTIMATE}) \times (100 - \text{ESTIMATE})$;
- if the other participant's decision was REJECT: $L = 0.01 \times \text{ESTIMATE} \times \text{ESTIMATE}$.

You can convince yourself that with this formula, you will earn an amount between £0.00 and £1.00, and that you will earn more money if your ESTIMATE is closer to indicating correctly the other participant's DECISION. It will therefore pay off for you to report a good guess. In fact, your expected earnings are maximal if you report truthfully what you think is the chance that the other participant ACCEPTED the proposed split. (We skip a more mathematical version of this property, and you can trust us on this. But fairly obviously, it has to do with the fact that L is a positive number, and that it is smaller the better is your estimate.)

Example: Suppose that the other participant ACCEPTS the proposed split choice (this decision is hypothetical, and not the actual decision of that other participant). Your task is to estimate this decision - you earn more points if your estimate better reflects the other participant's decision. With the above formula, you can verify that for this given outcome of the other participant's decision, you would receive:

- 100 points, if your estimate of the other participant's decision ACCEPTING the proposed split is 100%, or
- $100-9=91$ points, if your estimate of the other participant's decision ACCEPTING the proposed split is 70%, or
- $100-81=19$ points if your estimate of the other participant's decision ACCEPTING the proposed split is 10%.

If you have a question on this procedure, please raise your hand. Otherwise, please give us now your estimates that the other participant would ACCEPT each of the feasible splits (a percentage number between 0% and 100%).

If for some reason you want to change any of your decisions, simply re-enter a new number. You have to confirm your decisions (by clicking the OK button) to make them final. Once you confirm your decisions you will not be able to change them.

A.5 Screenshots

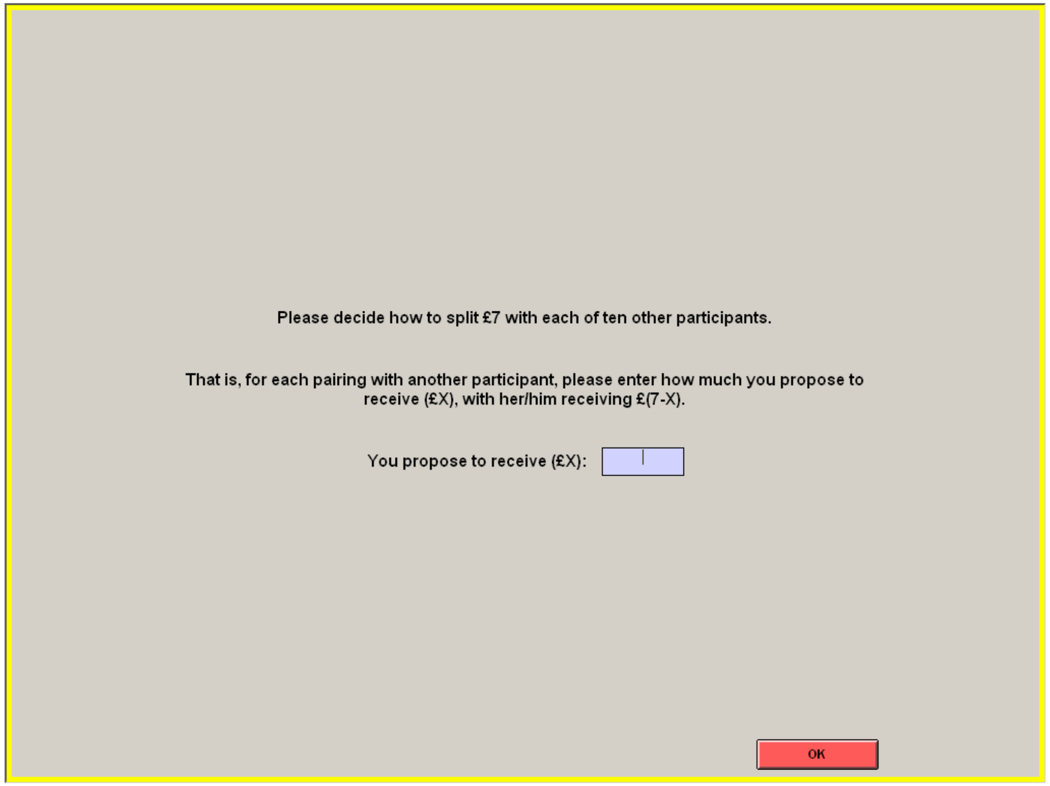


FIGURE A.1: Proposers' decision screen in both treatments

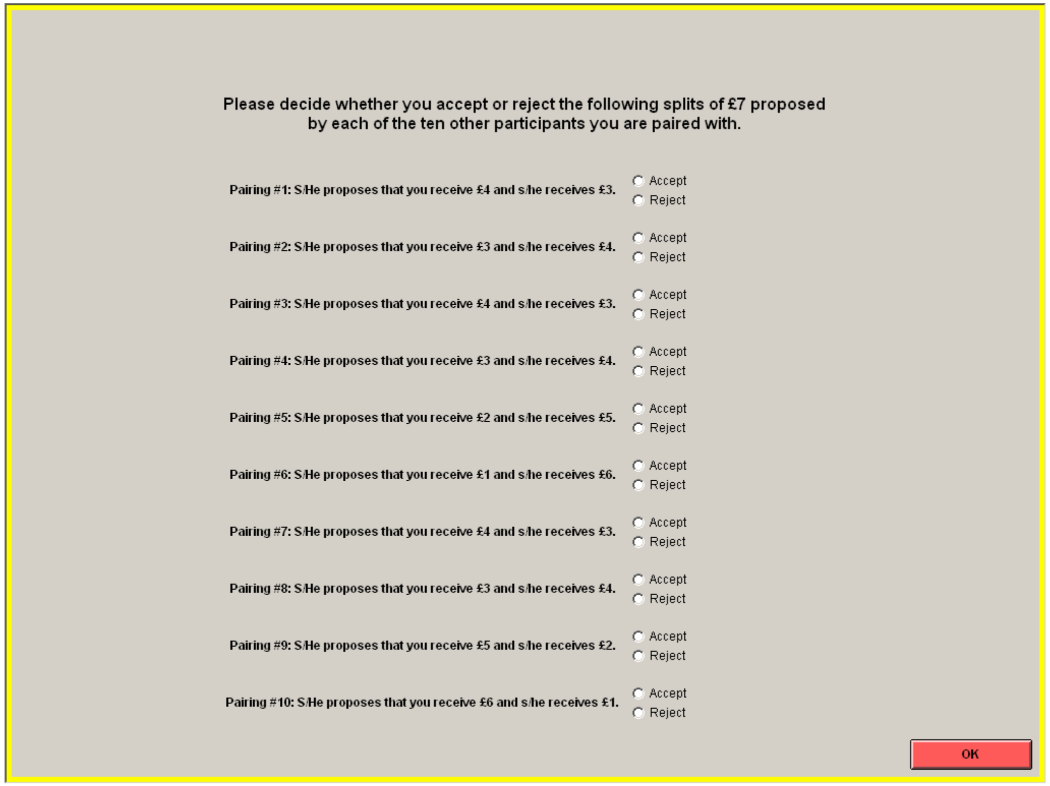


FIGURE A.2: Responders' decision screen in the response treatment

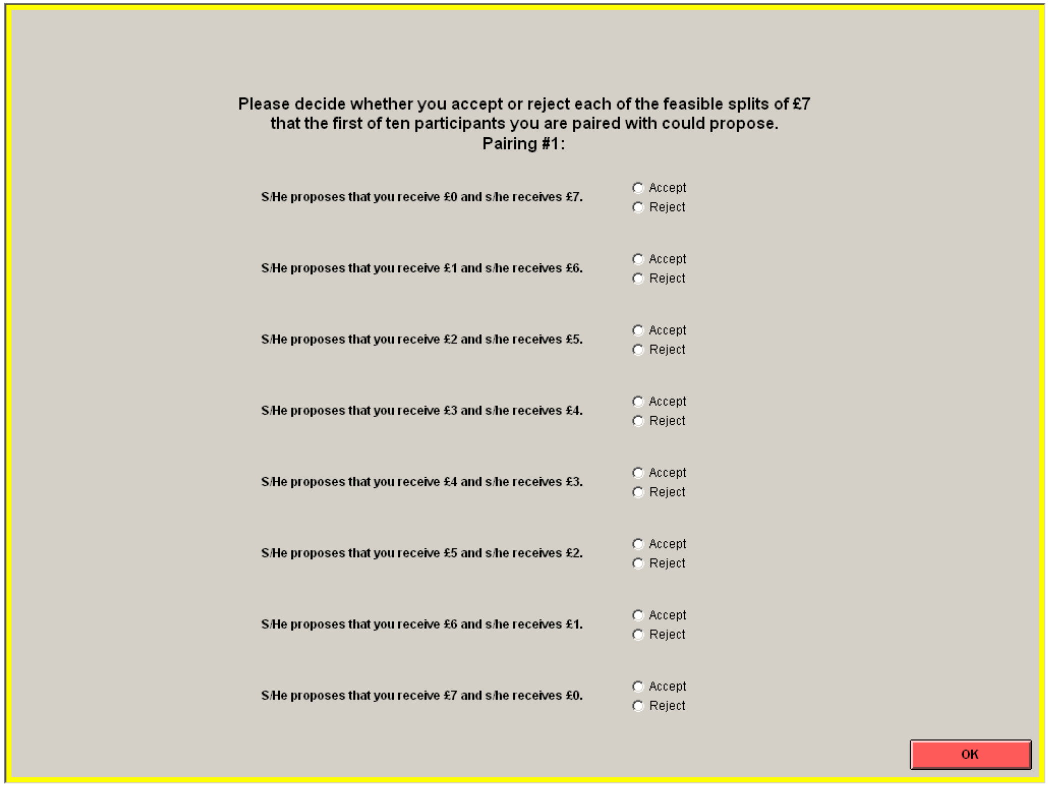


FIGURE A.3: Responders' initial decision screen for pairing #1 (optional 2nd screen for pairings #2 to #10) in the strategy treatment

FIGURE A.4: Responders' decision screen for pairings #2 to #10 in the strategy treatment

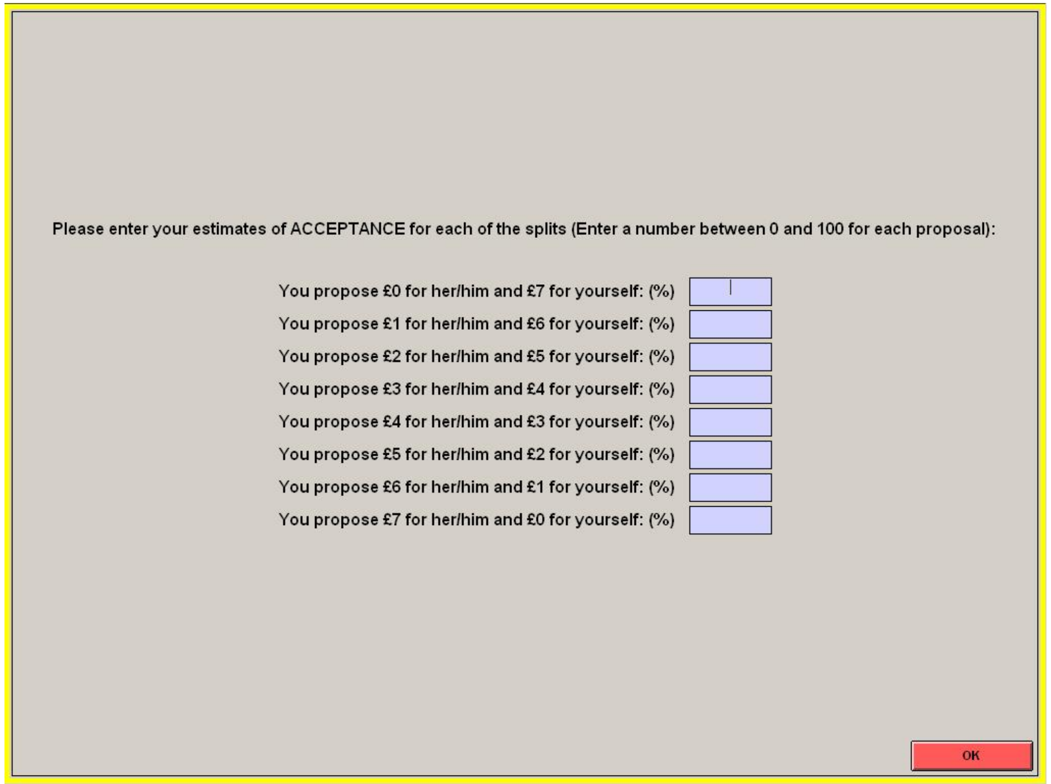


FIGURE A.5: Proposers' belief-elicitation decision screen in both treatments

Appendix B

Instructions and Screenshots for Experiment in Chapter 3

B.1 Instructions – Introductory remarks (common to both treatments)

PLEASE WAIT UNTIL THE EXPERIMENTER TELLS YOU TO START

Welcome to our experiment!

You are about to participate in an experiment in decision-making. In this experiment we will first ask you to read instructions that explain the decision scenarios you will be faced with. Next, you will be asked to make decisions that will allow you to earn money. Your monetary earnings will be determined by your decisions and the decisions of other participants who interact with you in the experiment. All that you earn is yours to keep, and will be paid to you in private, in cash, at the end of today's session. Your earnings will be kept strictly anonymous.

It is important that you remain silent and do not look at other people's work. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, exclaim out loud, etc., you will be asked to leave and you will forfeit your earnings. Thank you.

The experiment consists of three parts, parts I, II and III. You will participate in two parts. In each part in which you participate, you will anonymously interact with other participants in the room. You will not interact with any participant more than once throughout the experiment. Thus, if you interact with a participant in one part, you will not interact with him/her in the other part in which you participate. The decisions that you make in a part will influence NEITHER the decisions you will be faced with NOR the participants you will interact with in the next part in which you participate.

We will now randomly determine the two parts in which you will participate.

You will participate in parts <? > and <? >.

B.2 Instructions – No-Feedback Treatment

Instructions for trustors – Send stage

In this part, you will be anonymously paired with two other participants. For each pairing, you will be allocated £3 and the other participant will be allocated £0. Next, for each pairing, you will decide how much out of £3 to send to the other participant, keeping the rest for yourself. That is, for each of the two participants paired with you, you will send £X to him/her, and keep £(3-X) for yourself. Note that X has to be a whole amount in pounds between 0 and 3, i.e., (0, 1, 2, 3). The amount of money you send to each of the other participants, £X, will be tripled (by the experimenter). Hence, each of them will receive £(3×X). In summary, after you send £X to each of the two other participants, you will have £(3-X), and each of them will have £(3×X). However, the other participants will not be told how much you sent. Thus, neither of them will know how much s/he has as a result of your decision.

Next, each of the other two participants, will independently decide how much out of her/his money, £(3×X), to send back to you, with her/him keeping the rest for her/himself. Since s/he does not know how much s/he has, s/he will be asked to specify for each of the three feasible amounts s/he could receive (i.e., £3, £6, £9), how much s/he would like to send back to you, £Y which will also be a whole amount in pounds, and which can vary with the amount s/he has, £(3×X). (Note that the other participants do not need to specify how much you would return if you were to send £0, because they would receive £0, and so can only send you £0.)

After both you and they have decided, we will match your actual decision of how much you sent them, i.e., £X, with the decision each of them made on how much s/he would send back to you if s/he were to have £(3×X) after your decision, to determine the amounts earned by each of you.

Your earnings for each pairing will be equal to the sum of the amount you keep for yourself, i.e. £(3-X) plus the amount sent back to you by the other participant, £Y. Each of the other two participants will earn the amount they keep for themselves out of £(3×X) they have after your decision of sending £X, i.e., £(3×X-Y).

In summary, for each pairing, you will decide how much out of £3 you send to the other participant, keeping the rest for yourself. Each of the other participants receive the triple of the amount sent by you. They will separately and independently decide how much they would send back to you for each of your feasible decisions. After both you and they make your decisions, we will, for each pairing, match your decision with that of the other participant to determine that pairing's outcome, i.e., the amount earned by each of you. However, you will only be informed about the outcomes at the end of the entire experiment. At that time, one of the two pairings will be randomly chosen, and you will be paid the earnings of its outcome in cash. Thus, the chance that you will be paid for a particular outcome is one in two. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

*[Only shown in the **Feedback Treatment**]:* You will be informed about the actual amounts the others returned to you and your earnings at the end of this part.

Are there any questions? Please do not talk with others during the experiment.

Instructions for trustees – Return stage

In this part, you will be anonymously paired with two other participants. For each pairing, you will be allocated £0 and the other participants will be allocated £3. Each of the other two participants will independently decide how much of the £3 to send to you, keeping the rest for him/herself. That is, each of the other two participants paired with you will send £X to you, and will keep £(3-X) for her/himself. Note that X has to be a whole amount in pounds between 0 and 3, i.e., (0, 1, 2, 3). The amount of money each of the other participants sends to you, £X, will be tripled (by the experimenter). Hence, you will receive £(3×X). In summary, after each of the other participants sends £X to you, you will have £(3×X), and s/he will have £(3-X). However, you will not be told how much each of the other participants sent you. Thus, you won't know how much you have received for each of the pairings.

Next, you will decide, for each of the three feasible amounts you might receive (3×X)(i.e., £3, £6, £9), how much you would like to send back to each of the other two participants, £Y, keeping the rest for yourself. £Y is a whole amount in pounds, which can vary with the amount you have, £(3×X). (You do not need to specify how much you would return if you were to receive £0, because you could only send back £0.)

After both you and they have decided, we will match their actual decisions of how much they sent you, i.e., £X, with the decision you made on how much you would send back to each of them if you were to have received £(3×X) as a result of their decisions, to determine the amounts earned by each of you. Your earnings for each pairing will the amount you keep for yourself out of (3×X) (the money you have after they send you £X), i.e. £(3×X - Y). Each of the other two participants will earn the amount they keep for themselves out of their initial allocation of £3, £(3-X), plus the amount sent back by you £Y.

In summary, each of the other two participants will decide how much out of £3 to send to you, keeping the rest for him/herself. You will receive the triple sent by him/her. You will then decide how much you would send back to each of them for each of their feasible decisions. After both you and they make your decisions, we will, for each pairing, match your decision with that of the other participant to determine that pairing's outcome, i.e., the amount earned by each of you. However, you will only be informed about the outcomes at the end of the entire experiment. At that time, one of the two pairings will be randomly chosen, and you will be paid the earnings of that outcome in cash. Thus, the chance that you will be paid for a particular outcome is one in two. Likewise, for each pairing, the other participant also has the same chances (as you) of being paid his/her earnings.

*[Only shown in the **Feedback Treatment**]:* You will be informed about the actual amounts the others sent to you and your earnings at the end of this part.

Are there any questions? Please do not talk with others during the experiment.

Instructions for Idle-Subjects

You are not participating in this part.

Please be patient and wait until this part is over.

B.3 Instructions for Belief Elicitation stage (common to both treatments)

Instructions for trustors - After the Send stage

To finish this part, we now ask you to give us your ESTIMATE about the chance that one of the two participants paired with you would send back a certain amount out of what they received.

Specifically, given that that other participant has received $< £3 \times SEND >$, we ask you how likely you think that s/he would send back £0; how likely you think that s/he would send back £1; $\cdots$; how likely you think that s/he would send back $< \pounds 3 \times SEND >$.

You will be asked to answer with a percentage number. If you are absolutely sure that s/he would send back to you a specific amount out of $< \pounds 3 \times SEND >$, then you would want to answer with 100% for this amount. If you think it is absolutely certain that s/he would NOT send back a specific amount to you, then you would want to answer with 0% for that amount. If you are less certain that s/he would send back that amount, then you would want to respond with an intermediate percentage number (somewhere between 0% and 100%), reflecting what you think. Note that one and only one of the amounts that you are asked to state your estimate about will actually be sent back by her/him, thus the sum of all percentage numbers you stated should be 100%. Assigning a higher percentage number to a particular amount would indicate that you estimate a stronger tendency towards that amount being actually sent back by her/him, while assigning a lower number to it would indicate that you estimate a stronger tendency that the stated amount is not the amount actually sent back by her/him.

You will be rewarded for the accuracy of your estimates, as follows:

First, we select the participant whose decisions will be used to determine your earnings from your estimates. That participant is the one whose decisions were NOT randomly chosen to determine your earnings in the previous task, i.e., the one whose decisions of how much to sent back were NOT chosen to determine how much you earned given the amount of money you sent.

Next, we compute your reward which is equal to 100 points (each point is worth one pence), minus a number "L" (short for "loss") that indicates how well your estimate is when compared to the actual amount sent back by that other participant. You will be paid £0.02 for each point you earn in this task.

This number L is determined in several steps.

First, we look how well you estimated the amount actually sent back by her/him. We identify the decision of the other participant, i.e. look up how much he/she did send back to you. Your estimate (a percentage number between 0 and 100) for that particular amount will be compared with 100. We take the difference between 100 and your estimate (the percentage number between 0 and 100). The difference is squared (multiplied by itself), and then multiplied by 0.005.

Second, we also take into account how well you predicted the remaining possible amounts (which were not chosen by him/her) would be sent back to you. Each of these estimates (numbers between 0 and 100) will be squared (multiplied by itself) and multiplied by 0.005.

The number L is the sum of the numbers computed above.

You will earn an amount between £0.00 and £2.00, and you will earn more money if your estimate is closer to indicating correctly the amount sent back to you by her/him. It will therefore pay you to make a good guess. In fact, your expected earnings are maximal, if for each feasible amount, you report truthfully what you think is the chance that the other participant would send that amount back to you. (We skip a more mathematical version of this property, and you can trust us on this. But fairly obviously, it has to do with the fact that L is a positive number, and that it is smaller the better is your estimate.)

To illustrate what your payment would be for this task, we will consider three examples.

[For trustors who send £1]

Example 1: Suppose the other participant sends back £1, and your estimates for £0, £1, £2, £3 respectively are 10%, 20%, 30%, 40% . The number of points you would earn is $(100 - 0.005(0-10)^2 - 0.005(100-20)^2 - 0.005(0-30)^{-}0.005(0-40)^2 = 54)$. At £0.02 for each point, you will earn £1.08 for this task.

Example 2: Suppose the other participant sends back £2, and your estimates for £0, £1, £2, £3 respectively are 0%, 0%, 100%, 0%. The number of points you would earn is $(100 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(100-100)^2 - 0.005(0-0)^2 = 100)$. At £0.02 for each point, you will earn £2 for this task.

Example 3: Suppose the other participant sends back £3, and your estimates for £0, £1, £2, £3 respectively are 0%, 0%, 100%, 0%. The number of points you would earn is $(100 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(0-100)^2 - 0.005(100-0)^2 = 0)$. You earn nothing for this task.

[For trustors who send £2]

Example 1: Suppose the other participant sends £1, and your estimates for £0, £1, £2, £3, £4, £5, £6 respectively are 10%, 20%, 10%, 20%, 10%, 20%, 10%. The number of points you would earn is $(100 - 0.005(0-10)^2 - 0.005(100-20)^2 - 0.005(0-10)^2 - 0.005(0-20)^2 - 0.005(0-10)^2 - 0.005(0-20)^2 - 0.005(0-10)^2 = 62)$. At £0.02 for each point, you will earn £1.24 for this task.

Example 2: Suppose the other participant sends £2, and your estimates for £0, £1, £2, £3, £4, £5, £6 respectively are 0%, 0%, 100%, 0%, $\cdots$, 0%. The number of points you would earn is $(100 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(100-100)^2 - 0.005(0-0)^2 - \cdots - 0.005(0-0)^2 = 100)$. At £0.02 for each point, you will earn £2 for this task.

Example 3: Suppose the other participant sends £3, and your estimates for £0, £1, £2, £3, £4, £5, £6 respectively are 0%, 0%, 100%, 0%, $\cdots$, 0%. The number of points you would earn is $(100 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(0-100)^2 - 0.005(100-0)^2 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(0-0)^2 = 0)$. You earn nothing for this task.

[For trustors who send £3]

Example 1: Suppose the other participant sends £1, and your estimates for £0, £1, £2, $\cdots$, £9 respectively are 10%, 10%, $\cdots$, 10%. The number of points you would earn is $(100 - 0.005(0-10)^2 - 0.005(100-10)^2 - 0.005(0-10)^2 - \cdots - 0.005(0-10)^2 = 55)$. At £0.02 for each point, you will earn £1.1 for this task.

Example 2: Suppose the other participant sends £2, and your estimates for £0, £1, £2, $\cdots$, £9 respectively are 0%, 0%, 100%, 0%, $\cdots$, 0%. The number of points you would earn is $(100 - 0.005(0-0)^2 - 0.005(0-0)^2 - 0.005(100-100)^2 - 0.005(0-0)^2 - \cdots - 0.005(0-0)^2 = 100)$. At £0.02 for each point, you will earn £2 for this task.

Example 3: Suppose the other participant sends £3, and your estimates for £0, £1, £2, $\cdots$, £9 respectively are 0%, 0%, 100%, 0%, $\cdots$, 0%. The number of points you would earn is $(100-0.005(0-0)^2-0.005(0-0)^2-0.005(0-100)^2-0.005(100-0)^2-\cdots-0.005(0-0)^2=0)$. You earn nothing for this task.

N.B. The numbers used in these examples were selected arbitrarily. They are NOT intended to suggest how anyone might respond in any situation.

If you have a question about this procedure, please raise your hand. Otherwise, please give us your estimates (a percentage number between 0% and 100%) for each of the feasible amounts the other participant could return to you. If for some reason you want to change any of your estimates, simply re-enter a new number. You have to confirm your estimates (by clicking the OK button) to make them final. Your estimates will be accepted only if they add up to 100. Once you confirm your estimates you will not be able to change them.

A calculator will be available on the next screen where you enter your estimates.

Instructions for trustees - Before the Return stage

To finish this part, we now ask you to give us your ESTIMATE about the chance that one of the two participants paired with you would send you a certain amount out of the £3 initially allocated to her/him.

Specifically, we ask you how likely you think that that other participant would send you £0, how likely you think that the other participant would send you £1 (in which case you would receive £3 after £1 is tripled by the experimenter), how likely you think that s/he participant would send you £2 (in which case you would receive £6 after £2 is tripled by the experimenter), how likely you think that s/he would send you £3 (in which case you would receive £9 after £3 is tripled by the experimenter)?

You will be asked to answer with a percentage number. If you are absolutely sure that s/he would send you a specific amount, then you would want to answer with 100% for that amount. If you think it is absolutely certain that the other participant would NOT send that amount, then you would want to answer with 0%. If you are less certain that that other participant would send that amount, then you would want to respond with an intermediate percentage number between 0% and 100%, reflecting what you think. Note that one and only one of the amounts that you are asked to state your estimate about will actually be sent by her/him, thus the sum of all percentage numbers you stated should be 100%. Assigning a higher percentage number to a particular amount would indicate that you estimate a stronger tendency towards that amount being actually sent by her/him, while assigning a lower number to it would indicate that you estimate a stronger tendency that the stated amount is not the amount actually sent by her/him.

You will be rewarded for the accuracy of your estimates, as follows:

First, we select the participant whose decisions will be used to determine your earnings from your estimates. That participant is the one whose decisions were NOT randomly chosen to determine your earnings in the previous task, i.e., the one whose decisions were NOT chosen to determine how much you earned given the amount of money you sent back.

Next, we compute your reward which is equal to 100 points (each point is worth one pence), minus a number “L” (short for “loss”) that indicates how well your estimate is when compared to the actual amount sent by that other participant. You will be paid £0.02 for each point you earn in this task.

This number *L* is determined in several steps.

First, we look how well you estimated the decision actually made by her/him. We identify the decision of the other participant, i.e. look up how much he/she did send. Your estimate (a percentage number between 0 and 100) for that particular amount will be compared with 100. We take the difference between 100 and your estimate (the percentage number between 0 and 100). The difference is squared (multiplied by itself), and then multiplied by 0.005.

Second, we also take into account how well you predicted the remaining 3 possible amounts (which were not chosen by him/her) would be chosen. Each of these estimates (numbers between 0 and 100) will be squared (multiplied by itself) and multiplied by 0.005.

The number L is the sum of the 4 numbers computed above.

You will earn an amount between £0.00 and £2.00, and you will earn more money if your estimate is closer to indicating correctly the other participant's decision. It will therefore pay you to make a good guess. In fact, your expected earnings are maximal if you report truthfully what you think is the chance that the other participant would do. (We skip a more mathematical version of this property, and you can trust us on this. But fairly obviously, it has to do with the fact that L is a positive number, and that it is smaller the better is your estimate.)

Example 1: Suppose the other participant sends £1, and your estimates for £0, £1, £2, £3 respectively are 10%, 20%, 30%, 40%. The number of points you would earn is $(100-0.005(0-10)^2-0.005(100-20)^2-0.005(0-30)^2-0.005(0-40)^2=54)$. At £0.02 for each point, you will earn £1.08 for this task.

Example 2: Suppose the other participant sent £2, and your estimates for £0, £1, £2, £3 respectively are 0%, 0%, 100%, 0%. The number of points you would earn is $(100-0.005(0-0)^2-0.005(0-0)^2-0.005(100-100)^2-0.005(0-0)^2=100)$. At £0.02 for each point, you will earn £2 for this task.

Example 3: Suppose the other participant sent £3, and your estimates for £0, £1, £2, £3 respectively are 0%, 0%, 100%, 0%. The number of points you would earn is $(100-0.005(0-0)^2-0.005(0-0)^2-0.005(0-100)^2-0.005(100-0)^2=0)$. You earn nothing for this task.

N.B. The numbers used in these examples were selected arbitrarily. They are NOT intended to suggest how anyone might respond in any situation.

If you have a question about this procedure, please raise your hand. Otherwise, please give us your estimates (a percentage number between 0% and 100%) for each of the 4 feasible amounts the other participant could send you. If for some reason you want to change any of your estimates, simply re-enter a new number. You have to confirm your estimates (by clicking the OK button) to make them final. Your estimates will be accepted only if they add up to 100. Once you confirm your estimates you will not be able to change them.

A calculator will be available on the next screen where you enter your estimates.

B.4 Screenshots

FIGURE B.1: Trustors' decision screen in both treatments

FIGURE B.2: Trustees' decision screen in both treatments

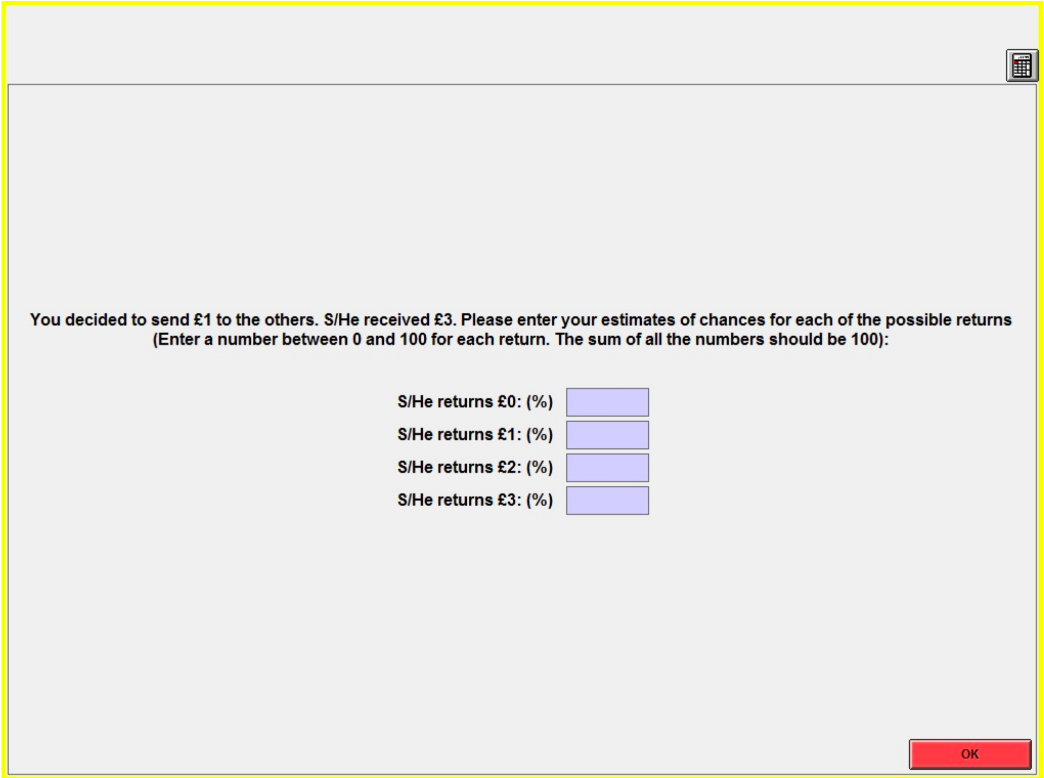


FIGURE B.3: Trustors' belief-elicitation decision screen in both treatments (for trustors who send £1)

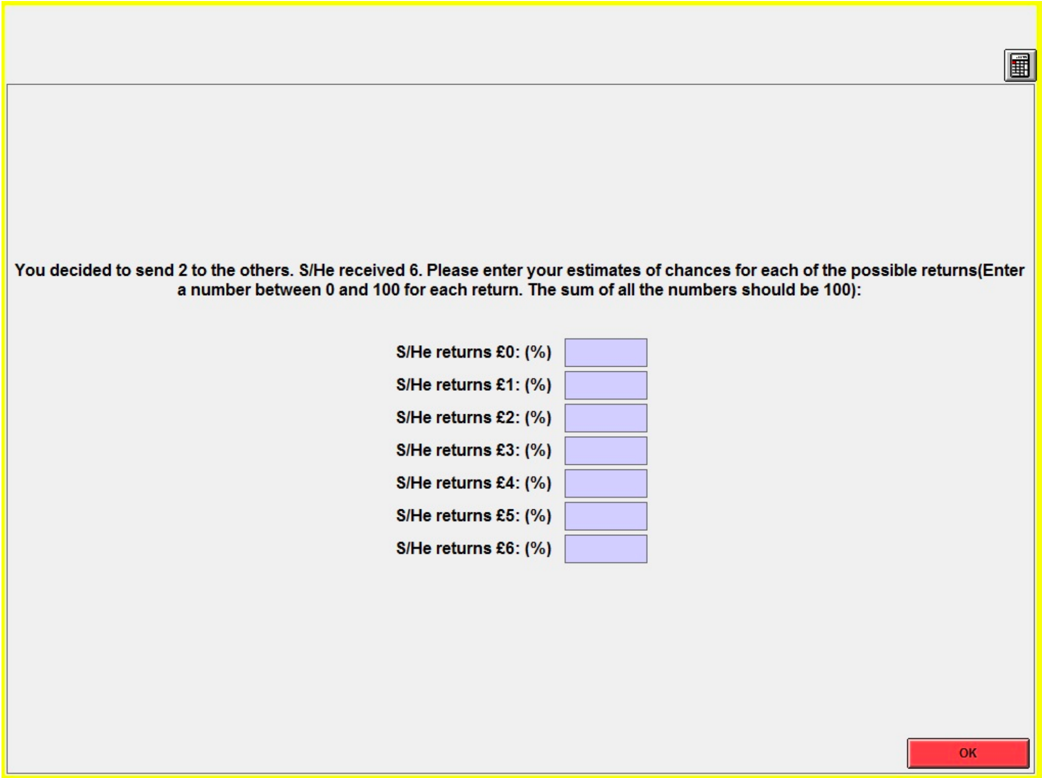


FIGURE B.4: Trustors' belief-elicitation decision screen in both treatments (for trustors who send £2)

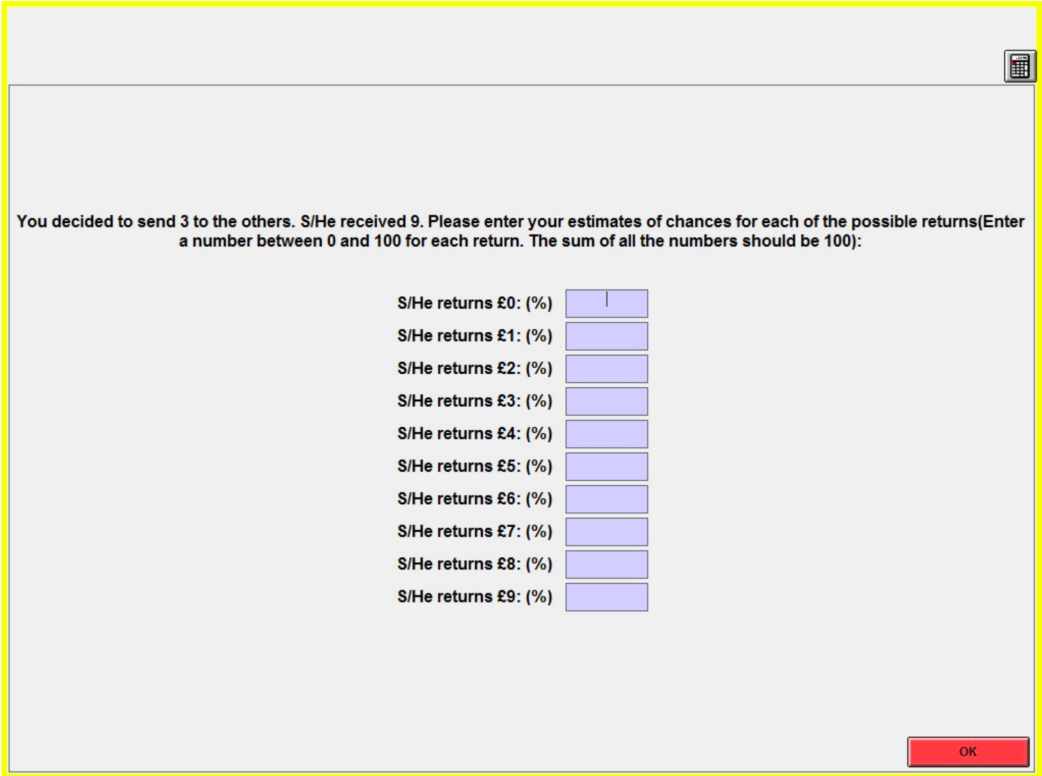


FIGURE B.5: Trustors' belief-elicitation decision screen in both treatments (for trustors who send £3)

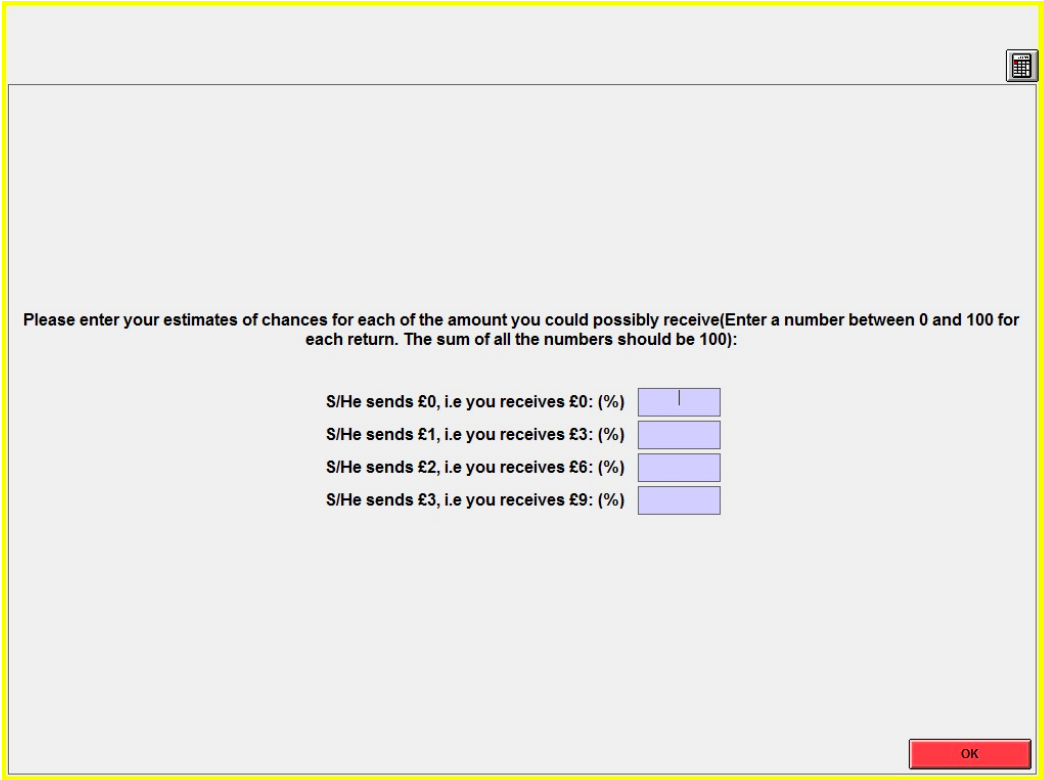


FIGURE B.6: Trustees' belief-elicitation decision screen in both treatments

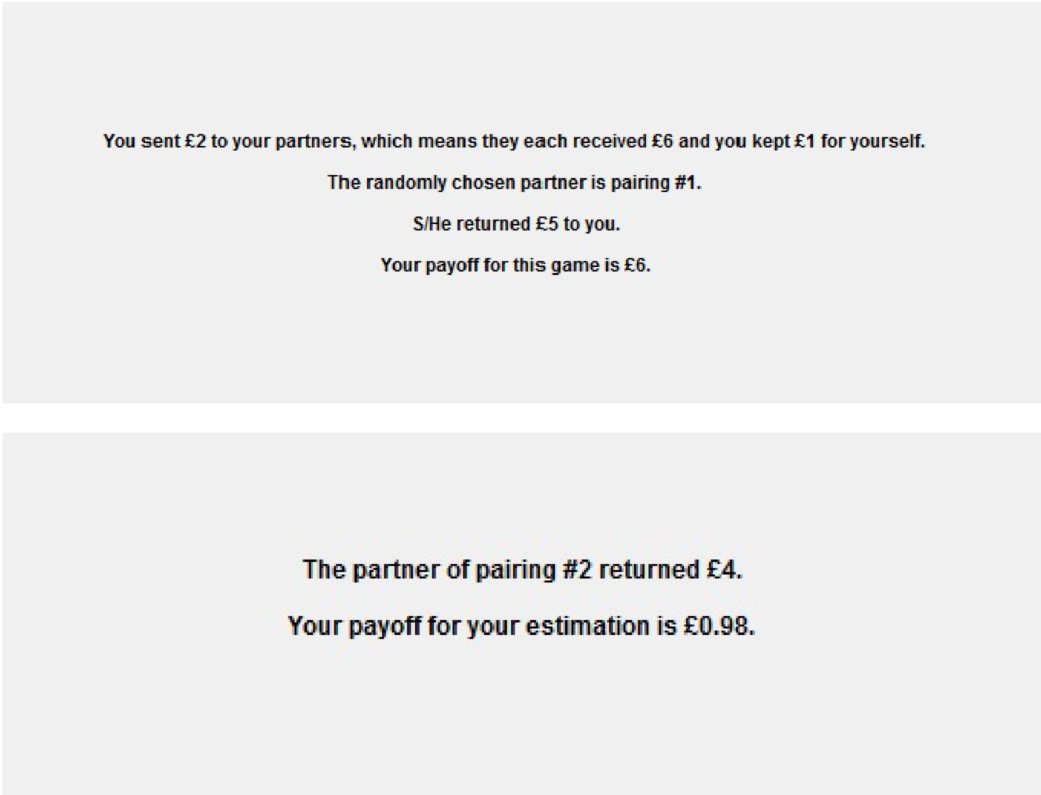


FIGURE B.7: Trustors' feedback in Feedback treatment

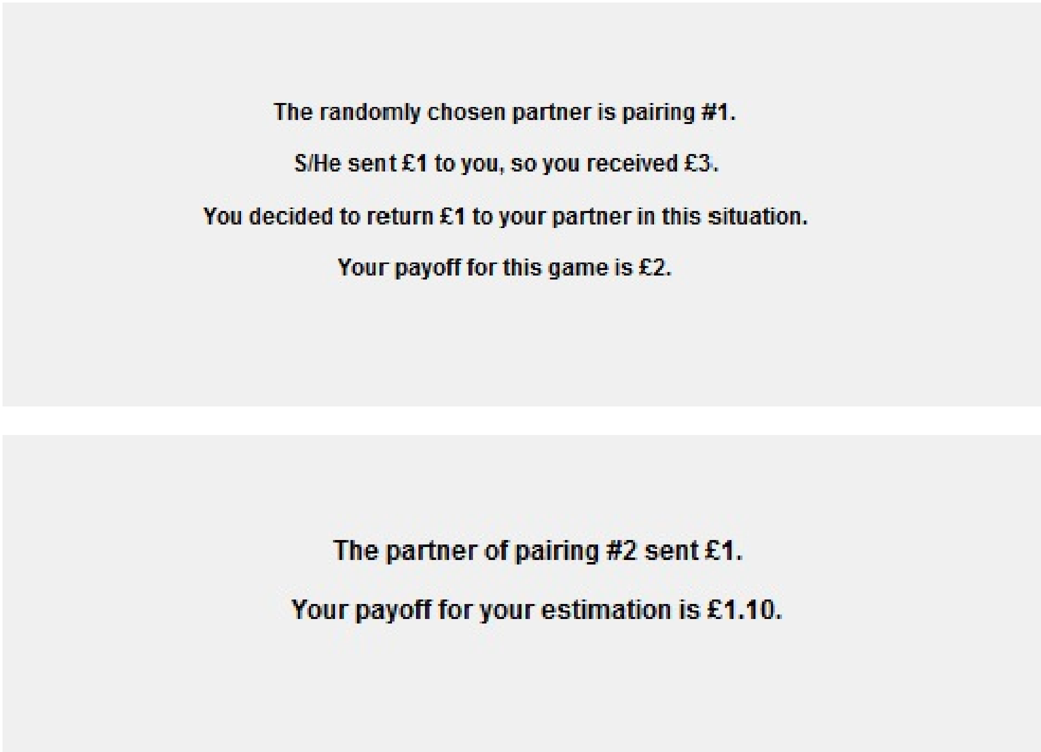


FIGURE B.8: Trustees' feedback in Feedback treatment

Appendix C

Instructions and Questionnaire for Experiment in Chapter 4

This appendix presents the English translation of the instructions[1] that we used in the Treatment PLU3SUB in which participants played three sequences of two games, and the post-experimental questionnaire. In the first two sequences, the interaction started at the implementation stage (and the participation stage was automatized by forcing everyone into the institution). In the third sequence the complete three-stage game was implemented as explained in the main text. Instructions for other sessions and the original Japanese versions are available upon request.

C.1 General Instructions

Welcome to our experiment!

You are about to participate in an experiment, which will help us to study decision-making and economic behavior. In this experiment, we will first ask you to read the instructions which explain the rules. Then you will be asked to make a series of decisions that will allow you to earn money. Your earnings will depend on the decisions you make and also on the decisions others make. During the experiment your earnings are counted in points. At the end of the experiment your earnings in points will be exchanged into Japanese yen at a rate of

[1] Instructions in Japanese and zTree codes are available upon request from the author.

2 Points = 1 JPY.

We will pay you at the end of the experiment in cash. Your identity, decisions and earnings will be kept strictly anonymous and confidential.

It is important that you remain silent and do not look at other people's work. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, exclaim out loud, etc., you will be asked to leave and you will forfeit your earnings.

There are 20 people participating in today's experiment. You will later be divided into two groups of 10 people. The experiment is divided into three sections. Each section consists of 20 rounds. You will interact with the same group of people in one section, but after each section, the members in your group will be changed. Other members of your group will face the same situation as yours. You will receive detailed instructions at the beginning of each section.

C.2 Instructions for Section I and II

In each round, you will receive 20 points, and you will be required to decide how many of them to keep to yourself and how many to contribute to a public project.

We first explain how the contribution you make to the public project benefit you and other members of the group. Your earnings in each round is the sum of the money you keep for yourself and the benefit from the public project, minus a cost. Expressed with the help of a formula:

1. Earnings $=$ points you keep $+$ benefit from the project $-$ cost

   The benefit from the public project is the same for each group member and it is determined in the following way:

2. Benefit from the project $= 0.4 *$ (total contribution by group members)

Now we explain how the value of cost is determined. The cost is of either 0 or 0.2 points, depending on the situation the group is facing. At the beginning of the game, you are automatically in an institution with all the other members of your group. The institution operates with a cost of 2 points that are shared by its 10 members. All members of this institution must vote in order to determine the level at which every member must contribute to the public project.

There are three options: "level 100%", "level 50%", and "level 0%". The option that gets the most votes will be chosen. In case of a tie, an option will be selected randomly among those that got the most votes. If "level 100%" is chosen, each member must contribute all 20 points and pay a cost of 0.2 points. If "level 50%" is chosen, each member must contribute half of his/her points (that is 10 points) to the public project and pay a cost of 0.2 points. If "level 0%" is chosen, the institution will be dissolved, which means there is no cost to be paid and all the members will be asked to decide on their own how many points to contribute to the public project.

Remember that earnings are calculated in the same way for each member of the group. So for each point you keep for yourself, you earn 1 point, but for each point you contribute to the project, each member of the group earns 0.4 points and the total earnings of the group increase by 4 points. Thus, your contribution to the project raises the earnings of everyone in the group. Other members' contribution increases your earnings in the same way. You earn 0.4 points for every extra point contributed to the public project by a members in the group.

You will receive information on your earnings at the end of each round.

Any questions? If you have a question during the experiment, please raise your hand.

C.3 Instruction for Section III

In each round, you will receive 20 points, and you will be required to decide how many of them to keep to yourself and how many to contribute to a public project.

Earnings are going to be calculated in the same way as in Section I and II. But in this section, at the beginning of each round, you will have the opportunity to decide whether to join the institution or not. If you join the institution, you must contribute your points according to the option that is most voted by the members of the institution, and share the cost with the other members who also joined the institution like in Section I. If you stay out of the institution, you can decide on your own how much to contribute to the public project.

After your decision, you will be informed about how many participants decided to join the institution. If there are less than 2 people who choose to join, the institution is not formed, which means no cost is to be paid and everyone in the group will decide on their own how much to contribute to the public project.

Any questions? If you have a question during the experiment, please raise your hand.

C.4 Post-experimental questionnaire

1. Age ______ Gender ______ Major ___________
2. Have you studied (or currently studying) Microeconomics?

   Yes ☐ No ☐
3. Have you studied (or currently studying) Game Theory?

   Yes ☐ No ☐
4. Have you studied (or currently studying) Economics?

   Yes ☐ No ☐
5. Have you ever heard of the Prisoners' Dilemma?

   Yes ☐ No ☐
6. Is your hometown located in any of the following major metropolitan areas? Tokyo, Nagoya, Osaka, Sapporo, Sendai, Yokohama, Kyoto, Kobe, Hiroshima, Fukuoka.

Yes ☐ No ☐

7. Do you live with your family?

   Yes ☐ No ☐

8. Do you consider yourself a cooperative person?

   Yes ☐ No ☐

9. Do you think that most people are usually cooperative?

   Yes ☐ No ☐

10. What do you think is the most efficient way to achieve a social goal?

    Cooperation ☐ Competition ☐

11. What was your goal in this experiment?

    Maximum payoff ☐ Maximum satisfaction ☐ Hurt the opponent ☐ Other ☐

12. Which of the following kind of associations (student circles) are you a member of?

    Sports (excluding gym) ☐ Cultural (music, theatre) ☐

    Environmental (Greenpeace) ☐ Other(please specify) ___________

13. How often do you use social networking websites, such as Facebook, Mixi, Twitter?

    Many times a day ☐ Normally once a day ☐

    Several times a month ☐ Almost never ☐

14. Have you ever taken advantages of someone?

    Yes ☐ No ☐

15. How much do you agree with the saying "good things happen to good people"?

    Strongly disagree ☐ Disagree ☐

Agree ☐ Strongly agree ☐

16. How much do you agree with the saying “no pain, no gain”?

    Strongly disagree ☐ Disagree ☐

    Agree ☐ Strongly agree ☐

17. To what extend do you think your opinion matters to the society?

    Very much ☐ Just slightly ☐ Not at all ☐

Bibliography

Alger, I. and J. Weibull. (2013). Homo Moralis: Preference evolution under incomplete information and assortative matching. *Econometrica*, 81: 2269-2302.

Allport, F.H. (1924). Social Psychology. Riverside Press, Cambridge, MA.

Altmann, S., T. Dohmen and M. Wibral. (2008). Do the reciprocal trust less? *Economic Letters*, 99: 454-457;

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

Ashraf, N., I. Bohnet and N. Piankov. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9: 193-208;

Atlamaz, M., C. Berden, H. Peters and D. Vermeulen. (2011). Non-cooperative Solutions for Estate Division Problems. *Games and Economic Behavior*, 73: 39-51.

Aumann, R. J. and M. Maschler. (1985). Game Theoretic Analysis of a Bankruptcy Problem from the Talmud. *Journal of Economic Theory*, 36: 195- 213.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46: 878-894.

Berg, J., J. Dickhaut and K. McCabe. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10: 122-142;

Bergstrom, T. (2009). Ethics, Evolution, and Games among

Neighbours. *Working Paper UC Santa Barbara*, available at: http://works.bepress.com/ted_bergstrom/106.

Bicchieri, C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms. *Cambridge University Press*.

Binmore, K., A. Shaked, and J. Sutton. (1985). Testing Noncooperative Bargaining Theory: A Preliminary Study. *The American Economic Review*, 75: 1178-1180.

Blackburn, S. (2001). *Ethics: A Very Short Introduction*. Oxford: Oxford University Press. p. 101. ISBN 978-0-19-280442-6.

Blanco, M., D. Engelmann, A. Koch and H.-T. Normann. (2010). Belief Elicitation in Experiments: Is there a Hedging Problem?. *Experimental Economics*, 13: 412-438.

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

Bolle, F. (1998). Rewarding trust: An experimental study. *Theory and Decision*, 45: 83-98;

Bowles, S. (2006). Microeconomics: Behavior, Institutions, and Evolution. Princeton University Press.

Brams, S. J., and A. D. Taylor. (1996). *Fair Division - From Cake-cutting to Dispute Resolution*, Cambridge University Press, Cambridge, UK.

Brandts, J. and G. Charness. (2011). The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons. *Experimental Economics*, 14: 375-398.

Buchan, N.R., R.T.A. Croson and S. Solnick. (2008). Trust and gender: an examination of behavior, biaes and beliefs in the investment game. *Journal of Economic Behavior and Organization*, 68: 466-476;

Burks, S.V., J.P. Carpenter and E. Verhoogen. (2003). Playing both roles in the trust game. *Journal of Economic Behavior and Organization*, 51: 195-216;

Burger, N.E. and C.D. Kolstad. (2009). Voluntary public goods provision, coalition formation, and uncertainty. *NBER Working Paper Series*, 15543, http://www.nber.org/papers/w15543.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in strategic interaction*. Princeton University Press.

Carraro, N. and D. Siniscalco. (1993). Strategies for international protection of the environment. *Journal of Public Economics*, 52: 309-328.

Chai, S.-K., D. Dolgosuren, M.S. Kim, M. Liu and K. Sherstyuk. (2011). Cultural Values and Behavior in Dictator, Ultimatum, and Trust Games. *Working Paper*. Availbale at: http://www2.hawaii.edu/ sunki/misc/research.htm.

Charness, G. and M. Rabin. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117: 817-869.

Chaudhuri, A. and L. Gangadharan. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, 73(4): 959-985;

Chun, Y. (1989). A Noncooperative Justification for Egalitarian Surplus Sharing. *Mathematical Social Sciences*, 17: 245-261.

Chun, Y. and J. Schummer and W. Thomson. (2001). Constrained Egalitarianism: A new Solution for Claims Problems. *Seoul Journal of Economics*, 14: 269-298.

Costa-Gomes, M. A., and G. Weizsäcker. (2008). Stated Beliefs and Play in Normal-Form Games.*Review of Economic Studies*, 75, 729-762.

Dagan, N. and R. Serrano and O. Volij. (1997). A Noncooperative View of Consistent Bankruptcy Rules. *Games and Economic Behavior*, 18: 55- 72.

Dannenberg, A. (2012). Coalition formation and voting in public goods games. *Strategic Behavior and the Environment*, 2(1): 83-105.

Dannenberg, A., A. Lange and B. Sturm. (2010). On the formation of coalitions to provide public goods–experimental evidence from the lab. *NBER working paper* no.15967.

Dannenberg, A., A. Lange and B. Sturm. (2014). Participation and commitment in voluntary coalitions to provide public goods. *Economica*, 81: 195-204.

Dufwenberg, M. and U. Gneezy. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30: 163-182;

Eckel, C.C. and R. K. Wilson. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization*, 55: 447-465;

Fehr, E., U. Fischbacher, B. Rosenbladt, J. Schupp, and C.G. Wagner. (2003). A nation-wide laboratory: Examining trust and trustworthiness by interating behavioral experiments into representative surveys. *IZA Disscussion Paper Series*, No.715;

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

Fehr, E. and U. Fischbacher. (2004). Social norms and human cooperation. *TRENDS in Cognitive Sciences*, Vol.8 No.4: 185-190.

Finus, M. and S. Maus. (2008). Modesty may pay. *Journal of Public Economic Theory*, 10: 801-826.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10: 171-178.

Garbarino, E., R. Slonim. (2009). The robustness of trust and reciprocity across a heterogeneous U.S. population. *Journal of Economic Behavior and Organization*, 69: 226-240;

Gerchak, Y., and J. D. Fuller. (1992). Optimal Value Declaration in "Buy-Sell" Situations, *Management Science* 38: 48-56.

Gleaser, E.L., D.I. Laibson, J.A. Scheinkman, and C.L. Soutter. (2000).

Measuring trust. *Quarterly Journal of Economics*, 115: 811-846;

Güth, W., S. Huck and W. Muller. (2001). The Relevance of Equal Splits in Ultimatum Games, *Games and Economic Behavior*, 37, 161-169.

Güth, W., H. Kliemt. (2010). What Ethics can Learn from Experimental Economics– If anything. *European Journal of Political Economy*, 26: 302-310.

Güth, W. and M.G. Kocher. (2014). More than thirty years of ultimatum bargaining: Motives, Variations, and a Survey of the Recent Literature. *Journal of Economic Behavior and Organization*, forthcoming.

Güth, W., R. Schmittberger, and B. Schwarze. (1982). An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization*, 3: 367-388.

Herrero, C. (2001). Equal Awards vs. Equal Losses: Duality in Bankruptcy. In M.R. Sertel and S. Koray, editors, *Advances in Economic Design*. Society for Economic Design, International Conference.

Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics*, 2: 141-159.

Hu, C.-C., M.-H. Tsay and C.-H. Yeh. (2012). Axiomatic and Strategic Justification for the Constrained Equal Benefits Rule in the Airport Problem. *Games and Economic Behavior*, 75: 185-197.

Isaac, R.M., K. McCue, and C. Plott. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, 26: 51-74.

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

Johansson-Stenman, O., M. Mahmud and P. Martinsson. (2011). Trust, trust games and stated trust: evidence from rural Bangladesh. *Journal of Economic Behavior and Organization*, forthcoming;

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

Kosfeld, M., A. Okada and A. Riedl. (2009). Institution formation in public goods games. *American Economic Review*, 99(4): 1335-1355.

Krueger, J.I. and M. Acevedo. (2005). Social Projection and the Psychology of Choice. In M.D. Alicke, J.I. Krueger and D.A. Dunning (Eds.) *The Self in Social Judgment*, Psychology Press, p15-p37.

Ledyard, J.O. (1995). Public goods. A survey of experimental research. In: J.H. Kagel and A.E. Roth (eds), *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press, 111-194.

Myerson, R.B. (1991). Game Theory: Analysis of Conflict. *Harvard University Press*.

Moulin, H. (2002). Axiomatic Cost and Surplus-Sharing. In: Arrow, K.J., Sen, A.K., Suzumura, K. (Eds.) *Handbook of Social Choice and Welfare*, vol.1: 290-357.

Ortmann, A., J. Fitzgerald and C. Boeing. (2000). Trust, reciprocity, and social history: a re-examination. *Experimental Economics*, 3: 81-100;

Oxoby, R. J. and K.N. McLeish. (2004). Sequential Decision and Strategy Methods in Ultimatum Bargaining: Evidence on the Strength of Other Regarding Behavior. *Economics Letters*, 84:399-405.

O'Neill, B. (1982). A Problem of Rights Arbitration from the Talmud. *Mathematical Social Sciences*, 2: 345- 371.

Pérez-Castrillo, D. and D. Wettstein. (2001). Bidding for the surplus: a non-cooperative approach to the Shapley value. *Journal of Economic Theory*, 100.2: 274-294.

Perry, M. and P.J. Reny. (1994). A noncooperative view of coalition formation and the core. *Econometrica*, 795-817.

Ross. L. (1977). The False Consensus Effect: An Egocentric Bias in

Social Perception and Attribution Processes. *Journal of Experimental Social Psychology*, 13: 279–301.

Roth, A. E. (1995). Bargaining Experiments. In John H. Kagel and Alvin E. Roth, editors, *The Handbook of Experimental Economics*, pages 253-348. Princeton University Press, Princeton, NJ.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara and S. Zamir. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81:1068-95.

Rubio, S. and A. Ulph. (2006). Self-enforcing international environmental agreements revisited. *Oxford Economic Papers*, 58: 233-263.

Sapienza, P., A. Toldra-Simats and L. Zingales. (2013). Understanding trust. *The Economic Journal*, doi: 10.1111/econj.12036;

Serrano, R. (1995). Strategic bargaining, surplus sharing problems and the nucleolus. *Journal of Mathematical Economics*, 24: 319-329.

Serrano, R. and R. Vohra. (1997). Non-Cooperative implementation of the core. *Social Choice and Welfare*, 14.4: 513-525.

Sonn, S. (1992). Sequential bargaining for bankruptcy problems. Mimeo.

Spooner, W.A. (1914). The Golden Rule. In J. Hastings, ed. *Encyclopedia of Religion and Ethics*, Vol 6, pp310-312.

Tan, J.H.W. and C. Vogel. (2008). Religion and trust: an experimental study. *Journal of Economic Psychology*, 29:832-848;

Thomson, W. (2003). Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences*, 45: 249-297.

Tu, Q. and E. Bulte. (2010). Trust, Market Participation and Economic Outcomes: Evidence from Rural China. *World Development*, 38: 1179-1190;

Walkowitz, G., C. Oberhammer and H. Henning-Schmidt. (2004). Experimenting over a long distance: A method to facilitate intercultural experiments. *Bonn Econ Discussion Papers*, No.2004, 17;

Wison, R.K. and C.C. Eckel. (2006). Judging a book by its cover: beauty and expectations in the trust game. *Political Research Quaterly*, 59: 189-202;

Zack, P.J. and S. Knack. (2001). Trust and Growth. *The Economic Journal*, Vol 111, pp295-321.