

Social Interactions and the Influence of “Extremists”

Ian Crawford*and Donna Harris†

July 9, 2018

Abstract

A large literature has shown evidence that people are influenced by others, especially in group interactions. However, little is known about whether such influence remains after they have left the group. Using a modified dictator game and a structural choice-revealed preference approach, we measure an individual’s preferences before and after face-to-face interactions in a small group and then examine whether a change in preferences is observed after subjects have left the group and have to make their decisions alone. We find that social interactions do indeed change individuals’ preferences. Specifically, individuals whose preferences are extremely egoistic and also unchanging tend to influence others the most. These “extremists” are more likely to be male and these effects are more prevalent amongst student subjects than non-student.

Key words: Other-regarding preferences, social interactions, preference dynamics, preference heterogeneity, social conformity

JEL classification: C90, C91, C92, D03, D71

*Email: ian.crawford@economics.ox.ac.uk. Department of Economics, University of Oxford and IFS.

†Email: donhatai.harris@economics.ox.ac.uk. Department of Economics, University of Oxford. The authors thank Jim Andreoni, Doug Bernheim, Martin Browning, Gary Charney, John A. List, Roberto Weber, three anonymous referees and the seminar participants at Stanford University, the University of Chicago and the University of Oxford for helpful comments and discussions. The authors also acknowledge the financial support from the Economic and Social Research Council (ESRC) and the Medical Research Council (MRC), Grant number: PTA-037-27-0187.

1 Introduction

The notion that “we are like chameleons, we take our hue and the color of our moral character, from those who are around us” is closely associated with the seventeenth century philosopher, John Locke.¹ In this paper, we experimentally study Locke’s observation in a particular context: the effects of social interactions on an individual’s preferences for generosity towards others. We find that Locke was right, social interactions do indeed change people’s behavior and preferences. However, we also find evidence that the effects of social interactions are not altogether positive: some subjects (especially students) may be encouraged to be more egoistic by other individuals who are already extremely selfish (the “extremists” of our title), even after they have left the group and have to make decisions again on their own.

Other regarding preferences² are an interesting subject of study in this context for a number of reasons. The first is that they determine how we treat others; social influence on preferences for private consumption by contrast, while interesting as well, do not. Secondly, whilst other-regarding behavior is often observed in various settings³, large heterogeneity is also observed across individuals (Benabou and Tirole, 2006; Fehr and Fischbacher, 2003; Henrich et al., 2005; Henrich et al., 2010; List, 2007)⁴. Moreover, research has shown that, just as our visual perception is influenced by the context, preferences (including other-regarding preferences) too are malleable and are affected by social environments such as elicitation method, the presentation (framing) of the problem, the ‘anchor’, and the salience of an aspect of one’s social identity, and inter-group conflicts (see Fehr and Hoff (2011) for a comprehensive review).

People can sometimes act differently in different situations and/or at different points in time, particularly when they are not alone. When people are able to observe, communicate and learn from each other or are being observed by others, they are likely to change their behavior to be more in line with those around them (Alpizar et al., 2008; Bardsley and Sausgruber, 2005; Cason and Mui, 1998; Charness and Sutter, 2012; Cooper and Kagel, 2005; Frey and Meier, 2004; Hoffman et al., 1996; Levitt and List, 2007; Luhan et al., 2009; Mas and Moretti, 2009; Shang and Croson, 2009; Soetevent, 2005; Vesterlund, 2003; 2006). There is, consequently, an emerging consensus that social interactions play a vital role in influencing individual behavior.

Social influence and social conformity have long been central to research in social psychology. One of the most well-known studies of social influence and social conformity was Solomon Asch (1952; 1956)’s line judgment experiment, in which subjects were placed in a group of confederates and were asked to estimate the geometric length of a line by matching it with one of three other lines. Each person in the group then took turns to publicly announce their answer. Despite the fact that the correct answer was obvious, Asch found that around a third of subjects gave the wrong answer when the confederates unanimously endorsed a clearly wrong answer. In an earlier study, Muiser Sherif (1936) showed that subjects conformed to the group norms even when they did not realize it and the effects also persisted afterwards, when they had to perform the same task alone without the presence of the group (unlike the Asch’s experiment in which the subject’s decisions were made in front of others). The literature of social conformity is vast and many theories have been developed in order

¹This is in fact a popular (mis)rendering; Locke wrote “We are all a sort of chameleons, that still take a tincture from things near us” in *Some Thoughts Concerning Education* (1693).

²Which can be defined broadly as the extent to which people care (either positively or negatively) about the material well-being of others around them (Cooper and Kagel, 2010). We prefer the term ‘other-regarding preferences’ to ‘social preferences’ because they can lead to either positive or negative behavior that are not necessarily ‘social’, such as the willingness to sabotage others at one’s own cost.

³For instance, people are willing to sacrifice a proportion of their own income to help others through charitable donations (Andreoni, 1989; 1990; 2006; Harbaugh, 1998; Karlan and List, 2007; Vesterlund, 2003); employees reciprocate positively (negatively) to firms that pay them above (below) the next best alternative by exerting higher (lower) effort (Falk, 2007; Gächter and Fehr, 2002); workers under a relative incentive pay scheme, where individual effort imposes a negative externality on others, exert less effort compared to those under piece rates because they do not want their friends to earn less (Bandiera et al., 2005).

⁴Stability of individual other-regarding behavior has been tested in both lab and field settings, but the results are inconclusive. Whilst some have found that other-regarding behavior observed in the lab is a good predictor of behavior in natural settings outside the lab (Benz and Meier, 2008; Carpenter and Seki, 2011; Carlsson et al., 2014; De Oliverira et al., 2012; Fehr and Leibbrandt, 2011; Rustagi et al., 2010; Stoop, 2014), others have found little correspondence of this behavior in the lab and the field (Carpenter and Myers, 2010; Karlan, 2005; Laury and Taylor, 2008; Stoop et al., 2012). Social interactions may help explain this heterogeneity.

to understand how and why people conform to the group norms (in the interest of space, we are not going to review them here, but see Cialdini and Goldstein (2004) and Rilling and Sanfey (2011) for extensive reviews).

In economics, traditionally research on decision-making focused on individual decisions and interactions were assumed to only occur through an anonymous process of price formation in the market. However, recently economists have become increasingly interested in understanding how non-market interactions affect individual behavior (Akerlof, 1980; Banerjee, 1992; Bernheim, 1994; Bikhchandani et al., 1992; Brock and Durlauf, 2000; Jones, 1984; Manski, 1993; 1995; 2000). The economic literature on social interactions now covers a wide range of contexts, including intra-household interactions (Becker, 1974; Bergstrom, 1989; McElroy, 1990; Rosenzweig and Wolpin, 1994), child custody outcomes in divorce proceedings (Flinn and Del Boca, 1995), crime and criminal networks (Glaeser, Sacerdote, and Scheinkman, 1996), college students experimenting with illegal drugs and sex (Duncan et al., 2005), peers influence on academic performance in high school (Gaviria and Raphael, 2001; Zimmerman, 2003) and in college (Sacerdote, 2001), and coordination on fertility practices (Kohler, 2001).

Similar to the social psychology literature, a large proportion of economic research on social interactions has focused on understanding how and why an individual's behavior varies with the behavior of the group. Many studies have demonstrated that people tend to adjust their behaviors to conform to what others do (see Manski (2000) for a review). There are a number of theoretical conjectures and experimental studies on why conformity is observed, such as limited private information that leads to herding behavior and informational cascades (Anderson and Holt, 1997; Banerjee, 1992; Bikhchandani et al., 1992; Eyster and Rabin, 2014; Huck and Oechssler, 2000; Hung and Plott, 2001); preferences for being like others or for maintaining a particular social status (Akerlof, 1980; Bernheim, 1994; Jones, 1984); and sanctions on deviations from prevalent norms (Bardsley and Sausgruber, 2005; Fehr and Gächter, 2000; Fischbacher, Gächter and Fehr, 2001; Fischbacher and Gächter, 2010). There are also increasing number of experimental studies of the effect of social interactions on individual behavior both in the lab (Cason and Mui, 1998; Eckel and Wilson, 2007; Knez and Camerer, 1995; Krupka and Weber, 2009) and in the field (Chen et al., 2010; Frey and Meier, 2004; Shang and Croson, 2009).

One way to study how social interactions affect behavior is to examine whether decisions made by a group are different from those made by an individual. Kluger et al., (2012) review existing literature in economics and psychology on differences between individuals and groups in interactive tasks (Prisoner's Dilemma, Ultimatum and Dictator games, Trust and other sequential games, and simultaneous games) and the results show that groups tend to behave more selfishly, seem to be more motivated by payoff maximization (although this is not always the case), and appear to be more competitive than individuals. In these studies, groups are required to make decisions as a single unit (although there is usually no explicit requirement on the decision rule that is used within the group to reach an agreement) and then group's decisions are compared to individual's decisions in the same context (in the same game).

Closely related to our paper are the studies of group decisions in Dictator games. Two conflicting results have emerged and currently the jury is still out. Cason and Mui (1997) reported a tendency of groups to be more generous in giving than individuals, whereas Luhan et al. (2009) found that groups are more selfish than individuals. According to Kluger et al. (2012), Luhan and colleagues argued that the differences in these findings may be due to the fact that different group sizes were used (two in Cason and Mui's study and three in Luhan et al.). The larger the group size, the more difficult for people outside the group (the recipient in this case) to identify who might be responsible for the selfish decision. This is known as 'identifiability' hypothesis, which proposes that in intergroup interactions responsibility for a choice is more obscured compared to interindividual interactions in which players are more easily identified and thus, can be held accountable if they make a competitive or selfish choice (Kluger et al., 2012). Another potential explanation for the differences in observed behavior is that Cason and Mui used a procedure where participants could be easily identified i.e. groups were called to the front of a main room to receive feedback and payment and were then excused to the hallway and the interaction was face-to-face discussion. In the study by Luhan and colleagues, on the other hand, everything was computer-based and thus, anonymity was preserved which made it more difficult to identify who was in the

same group.

Whilst in previous experimental economics studies the objective was to compare individual and group decisions, in our paper, we are interested in individual's decisions before and after social interactions and not the decision that is made by the group. Our experiment also differs from those in the social psychology literature, such as the Sherif's experiment which used tasks that were unrelated to other-regarding preferences and no monetary incentive was used. We, on the other hand, are interested in what happens to other-regarding preferences when people talk or deliberate with others whether to take an action, but then have to make the decision again *on their own* with real money at stake. For example, a person may contemplate donating money to an appeal but then they discuss with their friends who are opposed to the idea. We are interested in the following questions: will they subsequently make the same decision they originally did, or will they change their mind? More generally, do people become nicer or nastier once they have the opportunity to interact with others? In particular, we examine the interplay between heterogeneity of preferences across individuals and the influence of social interactions on the distribution of preferences. In other words, does the effect of social interactions vary with the type of preferences that a person has initially? Furthermore, our approach allows us to examine the extent to which exposure to individuals with extreme tastes/preferences may, or may not, rub off on the subsequent preferences and behavior of others.

Our Contributions

In this section, we outline our specific contributions to the literature. Our objective is to examine whether social interactions via face-to-face group deliberations change individual other-regarding preferences and, if so, to what extent does the change depend on one's own and other's preferences? Note that the question we want to address is neither (a) whether groups reach different decisions than individuals, nor (b) whether individuals behave differently while in a group.

Instead, we are interested in studying whether participation in group deliberations affects individuals' behaviors/preferences, particularly once they are away from the group and return to 'private' decision-making. In this case, 'identifiability' hypothesis cannot explain any change in an individual's behavior since he/she cannot hide behind the anonymous nature of group decision. Therefore, in our setting changes in behavior may be due to the individual internalizing some aspect of the group behavior or the deliberation. We believe that this question is pertinent to understanding many real world behaviors. For example, how socializing with the wrong group of friends at school can dramatically change one's behavior or how public debate can shape or change one's social attitudes.

Using a laboratory experiment and a structural choice-revealed preference approach which preserves heterogeneity across individuals, our study bridges the literature on individual-level analysis of other-regarding preferences⁵ and the literature on social interactions. In particular, our study contributes to the study of changes in other-regarding preferences as a result of group discussions (see Charness and Sutter, 2012 for a comprehensive review), but the key departure of our paper from other studies on group discussions and group decisions is that we are interested in the *ex post* effect of group discussions on individuals' preferences and have to make decisions without being witnessed by the other group members.

Furthermore, we study whether such effect varies between student and non-student subjects. The reason we think this is interesting is because there has been a number of papers which examine whether students behave differently from the general population (non-students), particularly with respect to other-regarding preferences,

⁵There is now a growing literature on individual-level analyses of other-regarding preferences in different experiments within the same sample. Andreoni and Miller (2002) were among the first to conduct an analysis of individual choices across several dictator games with different budget sets in order to examine whether subjects behaved consistently with the axioms of revealed preferences. Building on the results of Andreoni and Miller, Fisman et al. (2007) estimated individual utility functions from several two- and three-person dictator games and allowed for non-linear budget constraints. More recently, Blanco et al. (2011) analyzed and compared the performance of different models of other-regarding preferences both at aggregate and individual levels using variety of games rather than variants of the same game. They found that within-subject tests can differ markedly from aggregate-level analyses.

but the results are inconclusive. Whilst some studies find that students are more likely to behave more selfishly than non-students (Anderson et al., 2013; Belot, Duch, Miller, 2015; Carpenter, Burks, and Verhoogen, 2005), others find no difference between the two subject pools (Falk, Meier, and Zehnder, 2013; Gächter, Herrmann, and Thoni, 2004). More importantly, little is known about whether social interactions affect students and non-students differently, which is our contribution to the literature on subject pool differences.⁶

In order to achieve our research objectives, we need to first measure preferences for each individual subject and then to find a way to compare them. We use Selten (1991)’s measure of predictive success to compare alternative models of other-regarding preferences that have been widely cited in the literature. Then, using the model that best accounts for our data, we recover and examine other-regarding preferences before and after social interactions. We do this by fitting our model to each individual separately thus completely preserving the heterogeneity which we are concerned with. Where a change in preferences is observed, we investigate further whether preferences of the individuals within the same group, who interact face-to-face, converge (or become more homogenous) i.e. whether there is evidence of social conformity. In addition, we also consider *who* may drive this change within the group and what kind of preferences they may have.

We find several notable effects of interactions. The first is that interacting with others face-to-face affects the subjects’ other-regarding preferences and the effects remain even after the subjects have left the group. Secondly, face-to-face interactions make preferences more homogenous. Third, and perhaps most interesting of all, the changes in preferences often seem to be driven by a single key individual within the group whose own preferences are often both extreme and also impervious to others’. We study the preferences of these “extremists” and find that they are often strongly self-interested and have little regard for inequality or social efficiency. They also tend to be male. We also find that student subjects are more easily influenced by these individuals than non-student subjects. Our results may therefore have interesting implications, particularly with regards to peers influence in young adults.

The paper is structured as follows. The next section describes the experimental design in detail. Section 3 describes the data and choice behaviors. Section 4 explains how we use the choice-revealed preference method to recover individual preferences and section 5 reports the main results on preference dynamics. We conclude in section 6 with a discussion of how our results relate to the literature and some suggestions for future research.

2 Experimental Design

We use the within-subject design, in which subjects are required first, individually, to make a set of decisions (incentivized) in a modified dictator game (described below), then participate in a ‘group deliberation task’ (not incentivized) to reconsider those decisions as part of an exogenously formed group, and finally to make those decisions individually once again (incentivized). Our interest is in comparing the choices/preferences of each individual before and after the group deliberation/social interaction phase, and also in comparing the choices/preferences of each individual with those of the others in their group.

In the group deliberation phase, the assignment to groups (of four or five individuals) was entirely random. The subjects (only the senders in the dictator game) were instructed to discuss the same choices, which they just made individually, together and to make the decisions together as a group. We did not specify any decision rules that they needed to use to reach the group decisions and they were clearly told that this group task would not have an impact either on their own or the receiver’s payoff. The aim of this intervention was to ensure that they interacted in a way which was focused on, and relevant to, the task they had just performed as individuals. By making them reach a group decision, the aim was, thus, to try to ensure that their social interactions were “on task” and not primarily about the weather, or the prospects for Oxford United in the forthcoming football season, but rather about the relative merits of different allocation decisions in a modified dictator game. However, we

⁶We thank two anonymous referees for suggesting that we highlight this contribution.

did not incentivize these group decisions because we did not want to induce interdependent preferences within the group. For example, a decision which appears to be selfish might be motivated by altruistic preference to help the fellow group members. By removing the incentive in this stage and exclude the group decisions from the payoff consideration and the analysis, we are able to focus on the individual decisions before and after the group deliberation which is our main interest.

We also ran a separate group of subjects who faced the same set of decisions and made them the same number of times (three times) with the same incentive structure⁷, but these subjects did not take part in the group deliberation phase (we refer to this treatment as the ‘non-deliberation’ game). The aim was to try to understand background level of choice reversals/preference instabilities which might be observed in any case amongst subjects in the absence of any social interactions - essentially, the ‘non-deliberation game’ serves as a baseline/control for the main treatment in which subjects interact face-to-face. We made sure that the gap between each stage was roughly the same as in the deliberation treatment group and subjects were also told that the second round would not be considered for the payoff calculation. Admittedly, this might seem unnatural to the subjects but none of the subjects questioned this procedure.

2.1 The Choice Set

For the decision task, we used a modified dictator game⁸ adapted from Charness and Rabin (2002). However, instead of presenting each choice individually, we presented the choices in a multiple price list format as shown in Table 1. This is because we asked subjects to make the same choices three times and therefore, presenting one choice at a time would be very time-consuming. It could be argued that allowing the subjects to see all the choices at once might encourage them to ‘psychologically hedge’ their decisions across the choices. For example, if a subject feels that she has been nice in her earlier choices, she might feel ‘justified’ to be more selfish in her later choices. However, this process is possible even if the subjects were presented with one choice at a time. Furthermore, if this was the case, we should observe this pattern in our data, which we did not. Specifically, if subjects psychologically hedged their choices, the ‘switches’ (if any) would be likely to occur towards the end of the list, but this was not what we observed. The reason for choosing this type of task rather than the systematic maps as in Andreoni and Miller (2002) is for simplicity in presenting the choices to the subjects given the multiple rounds and the group discussion task. A more complex task would be more time consuming and might introduce confusion during the deliberation phase.

In each stage, the sender (the dictator) was given a menu of 30 decision pairs. Each pair consisted of an allocation to the sender and another to an anonymous receiver who was seated in a different room. The sender was asked to choose one allocation per each decision pair and thus, made 30 decisions in total in a given stage. In both deliberation and non-deliberation groups, only the senders made the decisions, whilst the receivers were completely passive. All choice pairs were presented in Experimental Monetary Unit (EMU) and each EMU was worth 1 pence.⁹ Each choice pair differed in the sender’s own payoff, the receiver’s payoff, the total (joint) payoff, and in the payoff distribution between the two players. The choice set provides a rich choice environment that enables subjects to display different types of motives of other-regarding preferences and to capture a wider range of preferences than the standard dictator game where a dictator is given a fixed sum of money (Charness

⁷We thank Jim Andreoni, Doug Bernheim and Martin Browning for suggesting that we run this additional group (non-deliberation).

⁸In a standard dictator game (Forsythe, et al., 1994), a dictator is given an endowment of m dollars to divide between himself (person 1) and another participant (the receiver, person 2), so that $x_1 + x_2 = m$ where x_1 is the amount allocated to himself and x_2 is the amount allocated to the receiver. There is no strategic interaction or reciprocal motive in this game. The receiver is completely passive such that she has no other choice but to accept the amount offered by the dictator. If the dictator cares only about his own material payoff, he should keep the whole endowment to himself and transfers nothing to the receiver. The dictator game has been the subject of several studies (Hoffman et al., 1996; Camerer, 2003; for a meta study, see Engel, 2011), and on average people tend to offer in the range between 10% and 28% of the endowment to the receiver, which suggests that they do care about another person’s payoff.

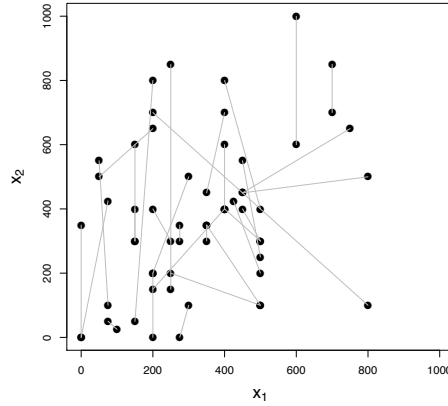
⁹We did not explicitly vary the “prices” as in Andreoni and Miller (2002). The relative price of self-payoff and the price of the receiver’s payoff was 1 as in the standard dictator game (and in budgets 7, 8, and 9 in Andreoni and Miller’s game).

and Rabin, 2002).

Table 1: The Choice Set

Choice	Either (x_1, x_2)	Or (x_1, x_2)	Choice	Either (x_1, x_2)	Or (x_1, x_2)
1	(400, 400)	(400, 600)	16	(400, 800)	(500, 400)
2	(0, 0)	(0, 350)	17	(500, 100)	(250, 200)
3	(600, 600)	(600, 1000)	18	(200, 700)	(800, 100)
4	(150, 300)	(150, 400)	19	(75, 50)	(100, 25)
5	(275, 300)	(275, 350)	20	(0, 0)	(75, 425)
6	(350, 350)	(350, 300)	21	(200, 200)	(300, 500)
7	(200, 200)	(200, 0)	22	(450, 450)	(750, 650)
8	(250, 850)	(250, 150)	23	(800, 500)	(450, 450)
9	(150, 600)	(150, 300)	24	(400, 400)	(200, 150)
10	(700, 850)	(700, 700)	25	(425, 425)	(500, 200)
11	(250, 300)	(200, 400)	26	(400, 400)	(500, 300)
12	(75, 100)	(50, 550)	27	(50, 500)	(200, 650)
13	(500, 100)	(350, 350)	28	(200, 800)	(150, 50)
14	(500, 250)	(450, 550)	29	(400, 700)	(350, 450)
15	(450, 400)	(500, 300)	30	(300, 100)	(275, 0)

Figure 1: The Choices in $\{x_1, x_2\}$ space



The set of 30 binary choices is illustrated graphically in Figure 1. The sender's own payoff (x_1) is on the horizontal axis and the receiver's (x_2) is on the vertical. Each binary choice is connected by a line. The two vertical lines furthest in the "northeasterly" direction, for example, refer to two binary choices $\{(600, 600) \text{ or } (600, 1000)\}$ which is choice number 3 in Table 1, and $\{(700, 850) \text{ or } (700, 700)\}$ ¹⁰ which is choice 10 in Table 1.

Choices 1 to 10 are such that the sender's payoff is always the same regardless of which allocation they select, but the amount allocated to the receiver and the total payoff vary. These choices allow us to look at the trade-off between social efficiency (total payoff) and relative payoff (inequality aversion).

Choices 11 to 19 allow the sender to sacrifice his own payoff in order to increase the receiver's income. These choices probe the altruistic motive i.e. the willingness to reduce one's own payoff in order to help another person. In addition, each of these choices also differs in the amount that the sender needs to sacrifice and in the receiver's payoff that is increased as a result. So, there is also efficiency concern in these choices. For example, in choice 11 the sender could sacrifice 50 EMUs to increase the receiver's payoff by 100 EMUs and thus, the

¹⁰This is similar to the game 'Berk29' in Charness and Rabin (2002).

sacrifice would result in ‘efficiency gain’ (total payoff increased from 550 to 600 EMUs). On the contrary, in choice 17 a sacrifice of 250 EMUs would only increase the receiver’s payoff by 100 EMUs, resulting in a loss in efficiency. In other words, being kind to the receiver is Pareto-damaging.

Choices 20 to 26 present the subjects with a trade-off between equal and unequal payoff distributions. We also vary own payoff and the total (joint) payoff. Specifically, choosing an equal distribution could yield either higher or lower payoff to the sender. Furthermore, in some of the choice pairs (20, 21, 22, and 23), an unequal distribution results in a higher total (joint) payoff (social efficiency gain), whilst in others (24 and 25) an unequal distribution results in a loss in efficiency and in choice 26, there is no loss or gain in social efficiency. The aim is to distinguish subjects who care about inequality from those who are concerned about social efficiency.

Finally, for *choices 27 to 30* one option allocates more money to both the sender and the receiver. We should observe the same choices being chosen by subjects who are selfish and by those who care about social efficiency since they both yield higher own payoff *as well as* higher total payoff. The senders who are inequality averse may choose differently, however. This is because we vary the payoff distributions in these choices (except choice 27). In addition, subjects who have jealous or spiteful preference may be willing to reduce the receiver’s payoff at their own cost (Andreoni and Miller, 2002) and thus, we may observe that some subjects choose the options that lower both players’ payoffs.

2.2 Experimental Procedure

The main treatment (‘deliberation game’) was carried out in April and July 2013, which were made of student subjects and non-student subjects. Student subjects were undergraduate students from University of Oxford and Oxford Brookes University who were randomly recruited via the Online Recruitment System for Economic Experiments (ORSEE) (Greiner, 2015) at Nuffield Centre for Experimental Social Science (CESS), University of Oxford. Non-student subjects were employees from an organization in London who signed up to attend a seminar given by Harris.¹¹ The subjects who participated in ‘non-deliberation’ game consisted of both student and non-student subjects recruited via ORSEE at CESS and was carried out in May 2015.

The experiment was carried out using pen and paper method for both deliberation and non-deliberation games. For the student subject pool, the experiment took place in two large seminar rooms at Nuffield College, University of Oxford and there were five sessions in total. Four sessions consisted of 30 subjects per session (15 senders per session) and one session consisted of 32 subjects (16 senders per session), which made up 76 sender decisions total. For the non-student subjects, we ran two sessions in two large seminar rooms with 50 subjects in the first session and 52 subjects in the second session (51 sender decisions in total). The organization has around 700 employees and thus, we recruited around 15% of the total population. The non-deliberation game was also conducted in the same large seminar rooms at Nuffield College with two sessions and 66 subjects (33 sender decision-makers in total, of whom 13 were students and 20 non-students). In total, there were 160 sender decisions and we took great care in making sure that the experimental procedure was exactly the same in all three groups and the same experimenter (Harris) and research assistants carried out all the sessions.

Upon arrival, subjects read and signed a consent form and randomly selected an ID number from an envelope provided by the experimenter. Those who picked the ‘odd’ numbers were randomly assigned to be the ‘senders’ (the dictators) and those who picked the ‘even’ numbers were assigned to be the receivers. The senders and the receivers were then directed to separate rooms. Only the senders made the decisions, whilst the receivers completed a long questionnaire that asked questions about inter-temporal choices, risk preferences, socio-economic background, and personality traits - all of which were un-incentivized.¹² But since we are only interested in the senders’ decisions, we do not report the results from the receivers’ questionnaires here.

¹¹The experiment was carried out prior to the seminar and as soon as the participants arrived and thus, their behaviors were not affected by the content of the seminar.

¹²Available upon request.

After everyone was seated, a hard-copy of the instruction was distributed and the subjects were given around ten minutes to read. Subjects were instructed to raise their hand if they had any question about the experiment and the experimenter would then go over to their seat and answered their question quietly. They were also told that no communication was allowed, unless instructed. The instruction was also summarized aloud by the experimenter to the group to ensure that everyone understood what they were asked to do. Both the senders and the receivers were provided with the same experimental protocol and thus, the procedure was common knowledge. Each sender (receiver) knew that he or she would be randomly paired with a receiver (sender), who was seated in a different room, and that the pair would remain *the same* throughout the experiment. However, they did not know the identity of their partner. The senders were clearly told that their decisions would determine both their own payoff and the receiver’s payoff and that the receiver’s task had no impact on the payoffs. Each receiver was also informed that his or her payoff would be determined by one of the sender’s decisions (whom he or she was paired with) and that this decision would be randomly chosen at the end of the experiment.¹³

Each stage of the decision task in the deliberation game was implemented as follows:

In *stage 1 (pre-deliberation individual choices)*, each sender was provided with a decision sheet consisting of the menu of 30 allocation pairs and were instructed to make decisions on their own without discussing with each other. We made sure that there was sufficient space between the seats, so that the answers could not be seen by other subjects (the seats were organized like an examination room). The experimenter then collected the decision sheets after everyone had completed it.

In *stage 2 (face-to-face group deliberation)*, we randomly grouped four or five *senders* together, depending on the number of people in the session. We used randomized group assignment¹⁴ rather than a naturally existing social groups, such as gender, ethnicity or nationality, because we wanted to avoid any influence of social identities on the subjects’ behavior. Of course, since we allowed people to see and talk to each other in a small group, their behaviors could still be affected by social identities of people within the same group. But since the groups consisted of different (randomized) compositions of gender, ethnic and nationality, it was less likely that a particular social identity could saliently drive the subjects’ behavior.

The randomized group assignment method worked as follows: once the senders had completed their decisions in stage 1, each subject (senders only) randomly selected a ‘group ID number’ from an envelope. An experimenter walked around the room and each subject picked a number at random one by one from the envelope. The group ID consisted of a letter and a number that represented the group that they were assigned to (the letter) and the member number (1-4 or 1-5). For example, subjects who randomly drew the group IDs ‘A1, A2, A3, and A4’ were all in group A; those who drew B1, B2, B3, and B4 were all in group B; and so on. Once all subjects had picked their group ID numbers, the experimenter instructed each group to move to a located area for the group discussion. Each area had a signpost with the group letter on it. In addition to the announcement by the experimenter, four research assistants were on hand to assist with the relocation to make sure that subjects did not talk to each other during the move. As a focus for the discussion, each group was given a ‘group decision sheet’ that was based on the same choice set as in stage 1 (but the order of the choices was randomized). They were then given 15 minutes to discuss and write down the group’s choices on one decision sheet. Their goal was to reach group decisions for all the 30 choices within 15 minutes.¹⁵ There were a total of 16 and 10 groups in

¹³The random pairing was done in Excel by the experimenter using the ID numbers. Subjects did not know who they were paired with.

¹⁴Our randomized group assignment is akin to the ‘Minimal Group Paradigm’, which has been shown to sufficiently induce the feeling of group affiliation, in-group favoritism and out-group discrimination in a large number of experiments in economics and social psychology (Tajfel, 1971; Tajfel and Turner, 1987; Chen and Li, 2009; Harris et al., 2015).

¹⁵We did not impose any sanction or punishment if the group did not reach an agreement. We simply told the subjects that they had 15 minutes to discuss and make group decisions. In all sessions, all groups reached an agreement without any problem. Moreover, since we wanted to keep the discussion as natural as possible, we did not assign any specific decision rules to the group. Each group was free to choose any decision rules they wanted in order to reach an agreement. The reason for asking them to make group decisions after the discussion was to give a focus to the discussion and to ensure that the subjects pay attention to the discussion topic, which was the decision task.

the student and non-student subject pools respectively. We did not exogenously group individuals according to their ‘type’ by using their first-stage individual decisions as in Luhan, Kocher, and Sutter, (2009). This is because we did not need to do this since we quantitatively measured preferences for *each individual* in the group based on their choices.

In *stage 3 (post-deliberation individual choices)*, the senders went back to their original seats (as in stage 1) and made individual decisions again based on the same choice set (the order of the choices was again randomized) without further communication. Note that the sender-receiver pairs remained *the same* through all three stages.

At the end of the experiment, the subjects (senders only) completed a short questionnaire which asked them about their socioeconomic background. Then *one* decision from either stages 1 or 3 (individual decisions only) was randomly chosen and was used to calculate the payments to both players. Therefore, only the individual decisions were incentivized and the subjects were clearly informed about this in the instruction (both in the deliberation and non-deliberation games).

In the non-deliberation game, all the procedures were the same, except for stage 2 in which subjects remained seated and were given the same decision task to complete alone. They were told that this stage would not be counted towards the payoff calculation but that they must complete the task in order to progress to the third stage.

In order to ensure transparency of the process, we randomly selected one of the subjects to pick the decision number (1 to 30) and the stage number (1 or 3) from an envelope. The selected decision was then publicly announced to both the senders and the receivers. The subjects were called out one-by-one to be paid in private. The average payoff was £8 including a show-up fee of £4 (in all three subject pools). The entire experiment lasted about 45 minutes in both deliberation and non-deliberation games.

Table 2: Summary of Subjects’ Characteristics

Variables		Non-deliberation	Deliberation	<i>p-value</i>	Students	Non-Students	<i>p-value</i>
Age	<i>mean</i>	32	28	<i>0.02</i>	25	33	<i>0.001</i>
	<i>std.dev</i>	13.14	7.71		7.34	9.63	
Male	<i>mean</i>	0.4	0.53	<i>0.17</i>	0.52	0.48	<i>0.30</i>
	<i>std dev</i>	0.50	0.50		0.50	0.50	
Undergrads	<i>mean</i>	0.53	0.40	<i>0.18</i>	0.35	0.53	<i>0.001</i>
	<i>std dev</i>	0.51	0.49		0.48	0.50	
UK	<i>mean</i>	0.64	0.56	<i>0.43</i>	0.32	0.90	<i>0.01</i>
	<i>std dev</i>	0.49	0.50		0.46	0.24	
Obs.		33	127		89	71	

Note: p-values are for a null of no-difference in the means.

The summary statistics of the subjects’ characteristics are shown in Table 2. In total, 160 subjects were assigned the role of the sender (127 in deliberation game, and 33 in non-deliberation game). Table 2 presents comparisons between treatments (deliberation vs. non-deliberation) and between subject pools (students vs. non-students). Comparing the non-deliberation to the deliberation groups, the only variable which was statistically significant is age. The average age of the subjects in the deliberation game (both students and non-students) was 28 years, whilst those in the non-deliberation game was higher (32) and showed a greater spread. But otherwise, the basic characteristics of the two groups were not statistically different and thus, in terms of the treatment and control groups our samples were quite balanced. For students and non-students, as to be expected students were younger than non-students (average age was 25 years compared to 33 years in the non-student pool). Students were more international (32% were British compared to 90% in non-student subject pool). But in terms of gender, there was no statistical difference between the two subject pools. It is important to note that the ‘undergraduate’ variable is slightly misleading here since this represents the proportion of non-student subjects who have at least completed an undergraduate degree. This does not include any

other professional qualifications such as accountancy and auditing qualifications which most of our non-student subjects have. This is why it looks like the proportion of undergraduate students is larger among non-students.

3 Choice Behavior

Next we examine the average choice behaviors. Table 3 presents the average amounts that the subjects allocated to themselves and to the receivers before and after group interactions. Similar to Table 2, we compare the deliberation and non-deliberation games; and student and non-student subjects. The p-values reported are for two-sample t-tests where the null is equality of means across the groups.

Table 3: Summary of Subjects' Choices

Amount to self	Non-delib.	Deliberation	<i>p-value</i>	Students	Non-students	<i>p-value</i>
Before	3.67	3.57	0.0008	3.60	3.53	0.0471
After	3.68	3.56	0.0002	3.61	3.49	0.0004
<i>p-value</i>	0.747	0.753		0.6119	0.2790	
Amount to other						
Before	4.17	4.11	0.4407	3.98	4.38	≈ 0
After	4.18	4.30	0.1207	4.15	4.51	≈ 0
<i>p-value</i>	0.8908	0.0046		0.009061	0.1147	

Note: H_0 equality of means.

The maximum amounts that the senders could allocate to themselves and to the receiver were £8 and £10 respectively and recall that the choice set was fixed such that the amounts that they could allocate were predetermined by the experimenter. First, we look at the non-deliberation and deliberation games, the average amount allocated to 'Self' (top left panel) was slightly higher in the non-deliberation game (3.67 compared to 3.57 in the first stage and 3.68 compared 3.56 in the third, with p-values against a null of "no difference" in both cases close to zero). Within each group, there was no change in the allocation to self between the phases (differences of 0.01 in each group with p-values around 0.75). In both groups the senders allocated more money to the receivers than to themselves initially (bottom left panel, 4.17 and 4.11 for non-deliberation and deliberation games respectively). However, whilst we see no change in the allocation to receivers by the non-deliberation group in the third ("after") phase, there was a moderate but statistically significant increase in allocation to the receivers for those in deliberation game (p-value of 0.0046).

Next we compare the choice behaviors of students and non-students, students allocated more money to themselves in the first stage (top right panel, 3.60 compared to 3.53 and the difference is statistically significant as shown in Table 3). However, the amounts allocated to themselves did not significantly change after group deliberation. Students and non-students allocated more money to the receivers than to themselves (bottom right panel), but non-students allocated more to the receiver initially (4.38 compared to 3.93 and this difference is statistically significant). This could be because they are employees of the same organization and this could promote in-group bias within this sample, which is not necessarily present in the student sample.¹⁶

Another way to look at changes in behavior is to look at the distributions of the changes in choices (or 'choice switching' from stages 1 and 3). We would expect experimental subjects who repeatedly answer the same rather demanding set of decisions to alter their choices somewhat between successive rounds, either out of error, tedium or perhaps as a result of further reflection on the problem. But our conjecture is that whilst we would expect a degree of this to occur in the non-deliberation group, choice-reversals may be expected to be

¹⁶We thank an anonymous referee for pointing this out.

more prevalent in the deliberation groups where subjects may experience all of these effects, but are additionally exposed to the potential added influence of social interactions.

Figure 2: Empirical CDFs of the changes in choices, by subject group

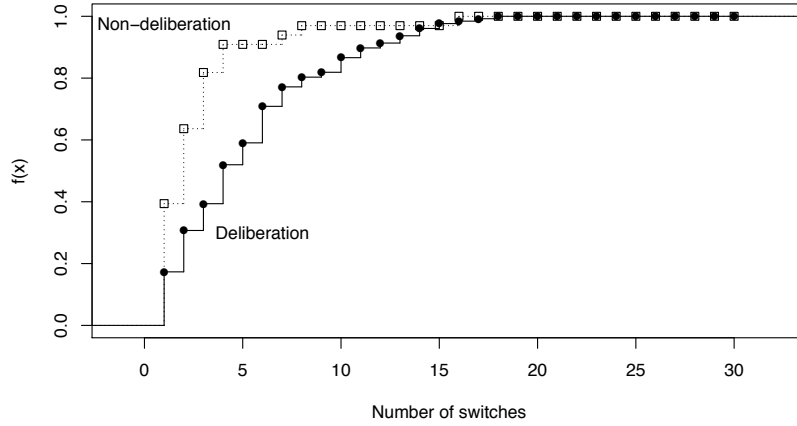


Figure 2 shows the distributions of switches/choice reversals for the treated (deliberation) and the control (non-deliberation) groups (a choice reversal occurs when, for example, a subject who chose the “either” in the first round switches to the alternative “or” option in the same choice pair in the second round). The horizontal axis records the number of choice reversals from zero (where subjects make exactly the same choices after the social interaction as they did before), to the maximum possible number of reversals 30 (in which a subject reverses every single one of the binary choices they previously made).

The main point of interest is to compare the rate of choice-reversals for the treatment (deliberation) group to the control (non-deliberation). As is evident from the distributions the deliberation group switched their choices much more than the non-deliberation group and indeed their distribution first order stochastically dominates the non-deliberation (the Kolmogorov-Smirnov statistic is 0.4245, and the p-value on a test of equality of distributions is 0.0002). From this we conclude that the treatment/deliberation was associated with a level of behavioral change above what might be expected of an experimental subject required to perform a task like this repeatedly.

4 Modeling Preferences

We are interested in determining whether and how individuals’ other-regarding preferences may have changed after the group interactions. We, therefore, assume that when individuals make their choices they are seeking to maximize well-defined preferences over the division of payoffs, and we assess five different functional forms of other-regarding preferences¹⁷ that incorporate either concerns for inequality, social efficiency, or maxi-min preferences into an individual’s utility function (in addition to self-interest) and examine which of these best explains our data. All of the models we consider are simple linear-in-parameters models in which the parameters are assumed to be weakly positive and are normalized to sum to one. These simple-to-interpret models allow the individuals to trade-off various features of the division of payoffs between themselves and the other player.

¹⁷Similar to Engelmann and Strobel (2004), we focus in the main on versions of the other-regarding preference models that have been frequently cited in the literature, namely Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002).

Our first model considers preferences over own payoff, the joint payoff and the variance in payoffs.

$$u(\mathbf{x}, \boldsymbol{\theta}_i) = \theta_{i,1}x_i + \theta_{i,2} \sum_{j=1,2} x_j - \theta_{i,3} \sum_{j=1,2} (x_j - \bar{x})^2 \quad (1)$$

This model is statistical in flavor - the subject is supposed to have preference over their own payoff, and moments of the distribution of payoffs (effectively the average payoff and its variance). We assume that all three parameters are positive and are normalized to sum to one. Thus, $\theta_{i,1} = 1$ implies the egoistic model, whereas $\theta_{i,2} = 1$ and $\theta_{i,3} = 1$ imply maximization of the joint (total) payoffs and minimization of inequality respectively.

The next is a two-parameter model which is of the type considered by Bolton and Ockenfels (2000). Bolton and Ockenfels proposed a simple model which captured three types of behavior observed in experiments: equity, reciprocity, and competition. One feature of this model is that people do not care about payoff differences in levels, but rather about equality between their share as a proportion of the total payoff. We implement a version of their model as

$$u(\mathbf{x}, \boldsymbol{\theta}_i) = \theta_{i,1}x_i - \theta_{i,2} \frac{1}{2} \left(\frac{x_i}{\sum_j x_j} - \frac{1}{2} \right)^2 \quad (2)$$

The two parameters $\theta_{i,1}$ and $\theta_{i,2}$ reflect preferences for the player's own payoff and inequality in payoffs, relative to 50:50.

Our third model is an instance of the class of models introduced in Charness and Rabin (2002). Charness and Rabin argued that heterogeneity in experimental subjects' behavior might be motivated by differences over their concern with social welfare - sacrificing to increase the payoffs for other subjects, especially those with the lowest payoff - rather than a concern with reducing differences in payoffs. In their simple linear, two-person model, the assumption is that in addition to own payoff, an individual's propensity to sacrifice for another person is characterized by three parameters: the weight on the other's payoff when she is ahead, the weight when she is behind, and the change in the weight when the other person has misbehaved (reciprocity). Their model, therefore, embeds difference (inequality) aversion, social-welfare preferences, and preference for own payoff. Since we use the dictator game in our study and thus, reciprocity plays no role in the decision-making process (the receiver is a passive player), we drop the last parameter in their model and use

$$u(\mathbf{x}, \delta_i, \lambda_i) = (1 - \lambda_i)x_i + \lambda_i \left(\delta_i \min_j \{x_{j=1,2}\} + (1 - \delta_i) \sum_{j=1,2} x_j \right)$$

where $\lambda_i \in [0, 1]$ measures how much a person cares about the other-regarding preference compared to their own, and $\delta_i \in (0, 1)$ is a parameter measuring the concern for helping the worst-off person versus maximizing the total social surplus. Setting $\lambda_i = 1$ corresponds to what Charness and Rabin describe as purely disinterested preferences, in which players care no more or less about their own payoffs than others'; and setting $\lambda_i = 0$ corresponds to the egoistic model. Setting $\delta_i = 1$ corresponds to judging the distribution of payoffs solely according to how well off the least well-off is; setting $\delta_i = 0$ corresponds to total-surplus maximization. In our empirical work, we implement the equivalent linear-in-parameters form

$$u(\mathbf{x}, \boldsymbol{\theta}_i) = \begin{cases} (1 - \theta_{i,1})x_i + \theta_{i,1}x_j, & x_i < x_j \\ (1 - \theta_{i,2})x_i + \theta_{i,2}x_j, & x_i \geq x_j \end{cases} \quad (3)$$

where $\theta_{i,1} = (1 - \delta_i)\theta_{i,2}$ and $\theta_{i,2} = (1 + \lambda_i [1 - \delta_i])^{-1}$.

We also consider variations on the model suggested in Fehr-Schmidt (1999). Fehr and Schmidt developed the theory of inequality aversion to explain heterogeneity in observed behavior in many laboratory experiments:

particularly the puzzle as why experimental subjects behave very selfishly in some games, whilst demonstrating rather fair behavior in other games. They showed that when concerns about equality is taken into account, this puzzle can be resolved. In their model, it is assumed that an individual cares about her own payoff and also dislike inequitable outcomes (in a material sense). She dislikes being either worse off or better off than other people and that she suffers more from inequality that is to her material disadvantage than from inequality that is to her advantage (Fehr and Schmidt, 1999). The form of the model is given by

$$u(\mathbf{x}, \boldsymbol{\theta}_i) = \theta_{i,1}x_i - \theta_{i,2} \sum_{j \neq i} \max\{0, x_j - x_i\} - \theta_{i,3} \sum_{j \neq i} \max\{0, x_i - x_j\} \quad (4)$$

where it is assumed that $\theta_{i,2} \geq \theta_{i,3}$. The parameter $\theta_{i,2}$ reflects the aversion to personally disadvantageous inequality, and $\theta_{i,3}$ reflects the aversion to inequality which is personally favorable. As well as the classic restriction that $\theta_{i,2} \geq \theta_{i,3}$, we also consider a version in which this is relaxed (denoted as model (4')).

Finally, we consider a four parameter extension of Fehr-Schmidt in which we add the possibility that the subjects might also have preferences for the joint payoff while at the same time relaxing the $\theta_{i,2} \geq \theta_{i,3}$ assumption.

$$u(\mathbf{x}, \boldsymbol{\theta}_i) = \theta_{i,1}x_i - \theta_{i,2} \sum_{j \neq i} \max\{0, x_j - x_i\} - \theta_{i,3} \sum_{j \neq i} \max\{0, x_i - x_j\} + \theta_{i,4} \sum_{j=1,2} x_j \quad (5)$$

4.1 Assessing the models

Assuming maximizing behavior the subject's 30 choices must simultaneously satisfy a system of 30 inequalities:

$$u(\mathbf{x}, \boldsymbol{\theta}_i) \geq u(\mathbf{x}', \boldsymbol{\theta}_i) \text{ if } \mathbf{x} \text{ is chosen over } \mathbf{x}' \quad (6)$$

(one for each of the 30 binary, "either \mathbf{x} or \mathbf{x}' ", choices the subject makes). The question of whether or not an individual is behaving consistently with a model is simply one of determining whether there exists a set of parameters such that these inequalities hold simultaneously: that is, such that the utility function assigns higher values to the chosen bundle than the rejected bundle in all of their observed decisions. Since all of the models are linear in parameters, determining this is a straightforward linear programming problem.

We proceed model-by-model and individual-by-individual (hence, preserving heterogeneity across individuals). For each subject and each model, we calculate how many of the subject's choices are jointly rationalizable. We do this for each of our five models. Thus, for example, all of an individual's choices may be perfectly rationalizable by our implementation of Charness and Rabin's model (if 30 out of 30 observations are jointly consistent with maximizing behavior given this form of preferences), but only 90% rationalizable (if 27 out of 30 observations are jointly consistent) by Fehr-Schmidt. We refer to this proportion of the data for an individual, which a model can explain, as its *pass rate*.

It would be natural to compare the performance of different models using the pass rate as the loss function. However, these models use different numbers of parameters and different restrictions on parameters values. Our version of the Bolton-Ockenfels model, for example, only has two parameters, whereas the extended Fehr-Schmidt model (model (5)) has four parameters and therefore should, *prima facie*, be expected to have a higher pass rate. Similarly, model (5) nests model (4') which itself nests the standard Fehr-Schmidt model (4). It is, therefore, *necessarily* the case that, in terms of agreement between theory and data, (5) must do at least as well as (4') which itself must (weakly) out-perform (4). Comparing these models only in terms of their pass rates is, therefore, unsatisfactory as more permissive/flexible models necessarily out-perform more parsimonious models in this respect.

In order to overcome this, we use a measure of "predictive success" proposed by Selten and Kriskcher (1983) and Selten (1991). This penalizes the pass rate with a measure of how flexible the model is. Selten's approach

calculates the subject’s pass rate (the proportion of their observed choices which are model-consistent) and compares this to proportion of *all possible choices* which the subject could have made that are model-consistent. Selten argues that as the proportion of possible actions which are model-consistent gets larger, it becomes easier for almost any choice behavior to be model-consistent as almost “anything goes”. In the end, if all possible actions are model-consistent then the model is unfalsifiable - there are no choices the subject could possibly make which could conflict with the theory.

Selten suggests using this measure of the proportion of possible actions that are model-consistent (which he calls the “area” and denotes by $a \in [0, 1]$) in order to penalize the pass rate (denoted by $r \in [0, 1]$) and to combine the two into an overall measure of predictive success: $m(r, a)$. Selten (1991) argues that empirically successful theories combine small values of a with large r : good pass rates for models which tightly constrain behavior are a sign of good predictive performance and equally, high pass rates for models which are so flexible that almost anything goes should not indicate a good predictive performance.¹⁸ He also provides an axiomatic argument that the trade-off between the ability to fit the data and the restrictiveness of the theory should be the difference measure¹⁹: $m = r - a$. The resulting index $m \in [-1, 1]$ can be viewed as a pass/fail indicator, corrected for the ability to detect rejections. The interpretation of m is as follows. As m approaches one, we have a situation in which the theoretical restrictions are extremely demanding, coupled with behavior which satisfies them: the sign of a successful model. As m approaches minus one, we have restrictions which allow almost any observed behavior and yet the actual behavior fails to conform: the sign of an almost pathologically bad model. As m approaches zero we know we have a situation in which the ability of the model to rationalize the data simply mirrors the proportion of choices which *could* be theory-consistent.

Table 4: Model Comparison

Model	(1) Statistical	(2) Bolton-Ockenfels	(3) Charness-Rabin	(4) Fehr-Schmidt	(4') Unrestricted Fehr-Schmidt	(5) Extended Fehr-Schmidt
r	0.9311 [0.0200]	0.8674 [0.0260]	0.9303 [0.0209]	0.8768 [0.0201]	0.9243 [0.0268]	0.9421 [0.0185]
a	0.6948 [0.0364]	0.6822 [0.0349]	0.7049 [0.0331]	0.7359 [0.0361]	0.7730 [0.0368]	0.7718 [0.0331]
$m = r - a$	0.2363 [0.0415]	0.1851 [0.0435]	0.2254 [0.0392]	0.1409 [0.0413]	0.1512 [0.0455]	0.1693 [0.0379]

Table 4 shows the pass rate, area and Selten Index for the five models considered (standard errors in square brackets). The top row of the table shows the pass-rates for each model. In terms of pure agreement with the data, the extended Fehr-Schmidt model (model (5)) performs the best. It can precisely account for 94% of the experimental data (combining all treatments and subject pools). The worst performing model is the Bolton-Ockenfels inspired model (model (2)). But, these results could hardly be otherwise - model 5 must necessarily out-perform models (4) and (4') and, given that it has four parameters compared to model (2)'s and model (3)'s two parameters, it is to be expected that it does somewhat better than either of them.

The next row measures the proportion of all possible actions which each model could possibly account for²⁰ and this duly shows why model (5) does so well: of all *possible* actions subjects could take, 77% are model-consistent and only 23% are not. Thus, when the experimental subjects are shown to be behaving consistently with this model, it is certainly due in part to its relative permissiveness (the set of possible behaviors inconsistent with the model is relatively small). The Bolton-Ockenfels model is shown to be less permissive.

The third line is thus the one of interest. This reports the Selten measure of predictive success: the pass rate adjusted for the permissiveness of the model. In it we see that model (1) emerges as the best model -

¹⁸Selten’s measure of predictive success implements the idea that “A good theory will successfully predict events that are difficult to predict” (Binmore and Shaked (2010)).

¹⁹In brief, Selten’s main requirements are; monotonicity $m(1, 0) > m(0, 1)$; equivalence of meaningless theories $m(0, 0) = m(1, 1)$ and the requirement that the aggregate performance at the mean (across subjects) is equal to the mean performance (across subjects) $m(\bar{r}, \bar{a}) = \bar{m}$. This last assumption is strong and responsible for the linearity of the resulting index.

²⁰In our experiment there are 2³⁰ possible actions (combinations of binary choices). We therefore calculate this area numerically using a Monte Carlo method .

narrowly out-performing model (3). It has a good level of agreement with the data (at 93%, almost as good as the extended Fehr-Schmidt model) but with fewer parameters available to help in the task of rationalizing choices, it is less permissive. Table 3 pools both the before- and the after-deliberation data. In principle the model which fits the data best could change between the before- and after-deliberation phases. However this is not the case: in terms of the Selten measure, this model out-performs the others within each sub-phase as well as overall.²¹ We conclude, therefore that this model performs better than the alternatives we consider and it is our preferred model for what follows.

Figure 3: Indifference curve maps generated by model (1)

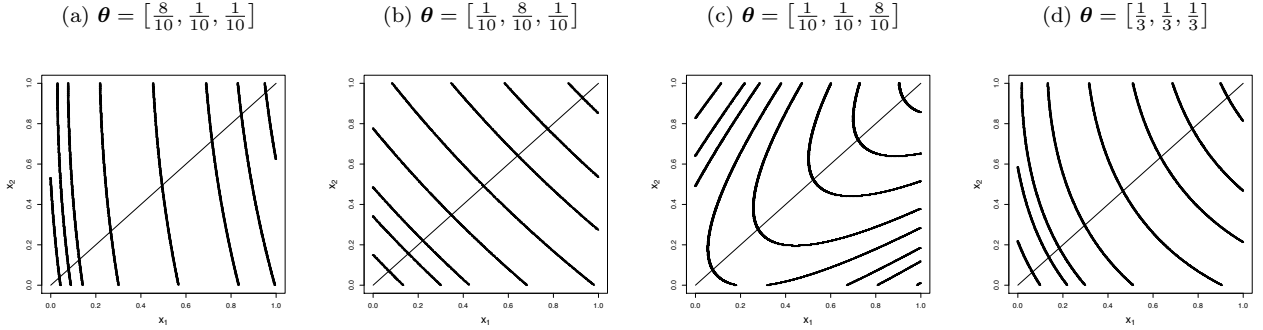


Figure 3 shows various indifference curve maps generated by this family of utility functions with a range of parameter values with $\theta = [\theta_1, \theta_2, \theta_3]$. More strongly ego-centric preferences tend towards vertical lines (3a), whereas regard for the total payoff (social efficiency) flattens the indifference curves (3b). Strongly inequality-averse preferences result in pronounced convexity (3c). Figure (3d) shows an equal weighting.

4.2 Recovering preferences

Given a set of model-consistent choices, the system of inequalities described by (6) and applied to model (1) identifies a convex *set* of parameters for each individual subject. Within this set (denoted S_i for individual i), we adopt the least-squares loss function and select the parameter vector which minimizes the integrated sum of squared errors with respect to all alternatives within the set.

$$\hat{\theta}_i = \arg \min_{\theta_i \in S_i} \int_{S_i} (\theta_i - \tilde{\theta}_i)^2 d\theta_i \quad (7)$$

We are able to compute both the rationalizing parameter set and the least-squares parameters for each subject in our data, before and after the group deliberation phase and compare them directly.

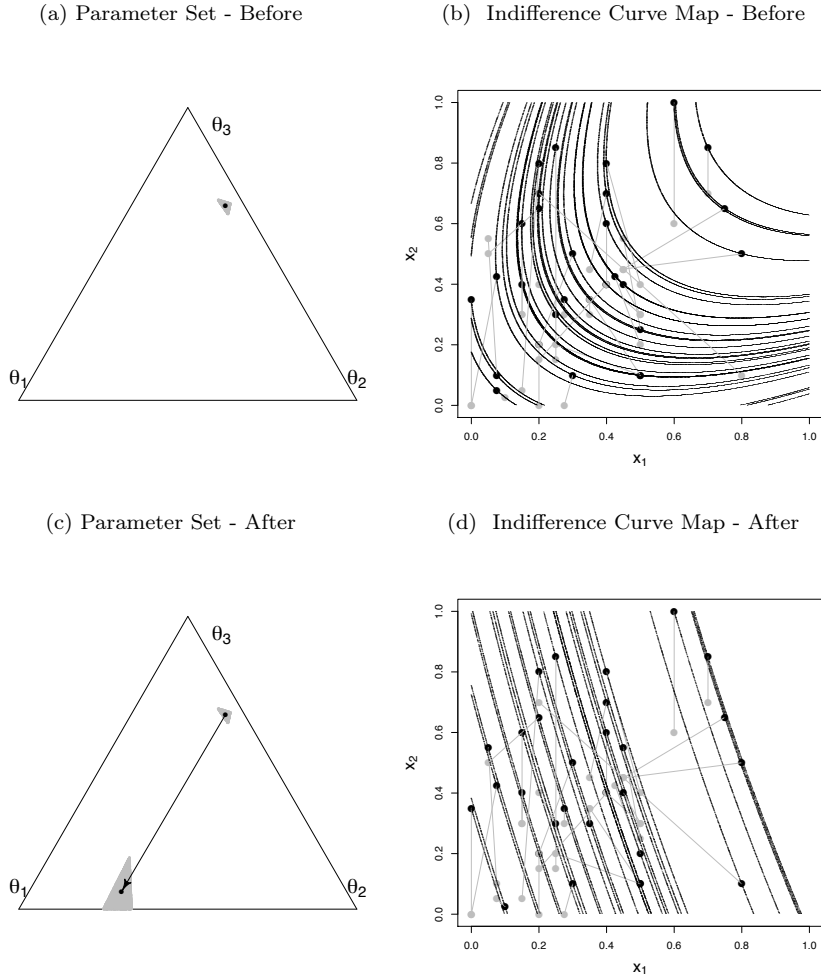
To illustrate and to fix ideas we have drawn one subject (#73) from the deliberation subject pool to show the parameter set, the least-squares parameter values and the corresponding indifference curves in each phase. *Before deliberation*, subject #73 selected the “Either” option in all choices in Table 1, *except* choices 3, 4, and 28. Afterwards this subject selected the “Either” option in all choices, except choices 22, 25, and 27.

Figure 4(a) shows the set of rationalizing parameters for this individual’s initial choices. Since our three parameters are positive and sum to one, we can represent the parameter space by the unit simplex and a particular parameter vector by a point within it. A completely self-interested subject would have preferences located at $\theta_1 = 1$ (bottom left), whereas a subject focussed on the joint pay-off would be at $\theta_2 = 1$ (bottom right) and a completely inequality-averse subject would have $\theta_3 = 1$ (top). The grey shaded polygon is the set of rationalizing parameters for this subject’s initial choices and this corresponds to S_i in the notation above.

²¹Results for each sub-phase are available from the authors. We are grateful to an anonymous referee for suggesting this comparison.

All points in the parameter set relate to parameters of an indifference curve maps which can rationalize this subject's choices. The marker in the centre of the parameter set shown in Figure 4(a) represents the preference parameters identified by our least squares loss function. The indifference curve map corresponding to these parameter values is illustrated in Figure 4(b). As shown, the rationalizing preference map always has indifference curves which pass through the individual's choices (black dots) and above the rejected alternatives (grey dots). As we can see from Figure 4a and 4b, this subject's choice-revealed preferences placed relatively little weight on their own pay-offs, but a lot of weight on inequality-aversion with the result that this individual has a very convex indifference curve map.

Figure 4: Subject #73



We then observed this individual's choices *after the group deliberation* stage of the experiment. Now subject #73 selected the "Either" option in all choices, except choices 22, 25, and 27. Since their choices altered, the preferences revealed by those choices also altered and this subject's new preference parameter set is illustrated in Figure 4(c). This subject, who initially was rather self-less and inequality-averse became much more selfish and their concern for inequality dropped away almost entirely (indeed as can be seen from Figure 4(c) there are parameter values which rationalize their behavior with $\theta_3 = 0$).

The indifference curve map corresponding to the least-squares parameters is shown in Figure 4(d). These indifference curves are close to linear, reflecting the newly acquired insensitivity to inequality and the preference for the aggregate payoff. They are also "steep" with respect to the individual's own payoff which indicates the greater weight the subject now places on this feature of the distribution. Comparing the parameters values (Figure 4(a) and Figure 4(c)) and/or the indifference curve maps (Figure 4(b) and Figure 4(d)) before and after

the group deliberation stage shows us how this individual’s preferences have changed. The Euclidean distance between before and after parameter values then provides a natural metric for preference change within our parameter space.

Our approach is therefore somewhat different to that of, for example, Fisman, Kariv and Markovits (2007) who also fit a structural model of preferences to experimental data. In their particular case subjects made continuous choices on budget constraints and therefore a continuous demand-system type approach (based on the CES in their case) is appropriate, as their experiment consciously “does not ‘force’ subjects into discrete choices” (Fisman, Kariv and Markovits (2007, p1861)). Our experiment, by contrast, only allows a sequence of binary choices. Some kind of discrete choice model would, therefore, seem a natural alternative approach. The difficulty with those models for the present application is that they explain departures from average behavior by random parameters. The joint distribution of these random parameters is pre-specified (e.g. they are often assumed to be jointly normally distributed with zero mean and uncorrelated, or follow an extreme value distribution etc.). However, the patterns in preference heterogeneity and in preference dynamics was something that we wanted to discover from the data rather than to constrain *ex ante*. Furthermore, with our set-up, preference parameters can only be partially (set) identified by the inequality restrictions implied by maximizing behavior in (6). The combination of multiple discrete choices, the wish to avoid assuming a distribution for the preference heterogeneity and partial identification together motivates why we choose our approach. The drawback is that if some of an individual’s choices cannot be rationalized by our model of preferences (since our model is deterministic, it does not include an error term to “explain” deviations from the model) then those observations are dropped from the analysis for that subject. This is controversial. Precisely how controversial depends on how much of the data is being dropped and whether there is something special/systematic about it. For our preferred model, we are able to retain 93% of the data (see Table 4), thus we typically cannot explain and therefore lose between 1 and 3 observations out of 30 for each subject. In Appendix A, we analyze the error rates and describe the excluded choices in detail. To summarize: the error rates (under our preferred model) are very low and the choices to which these errors were confined were essentially those where the alternatives were very close and possibly therefore hard for subjects to distinguish and consistent with small errors (the either/or choice between (350, 350) and (350, 300) is shown to be most often the source of the inconsistencies we observe).

5 Analysis of Preferences

We first briefly look at the average treatment effects of deliberation/non-deliberation. We then move to our main focus in this study: the effects of social interactions on preference heterogeneity across individuals and preference dynamics within each group.

5.1 Average Preference Effects

In order to describe the results we regress the individual preference parameters $(\theta_1, \theta_2, \theta_3)$ on the treatment variable (deliberation = 1, non-deliberation = 0), a dummy for third-stage preference (after deliberation/contemplation = 1, before deliberation/contemplation = 0), subject pool dummy (students = 1, non-students = 0), plus a number of covariates including gender (male = 1); age; education level (undergraduate = 1, masters and PhD = 0), and whether subjects were British or not (British = 1 and 0 otherwise). We also include interaction terms between the treatment (deliberation), third stage dummy, subject pools, and the covariates in some specifications of the regression (not all reported here but available upon request). All standard errors reported are clustered at individual level and the results for various specifications are all presented in Table 5²²

²²We thank an anonymous referee for this suggestion.

Table 5: Preference Parameters: regression analysis

	Reg (1)			Reg (2)			Reg (3)			Reg (4)			Reg (5)		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Deliberation (treatment)	-0.07 [0.05]	0.03 [0.03]	0.04 [0.04]	-0.11* [0.06]	0.013 [0.03]	0.10** [0.05]	-0.15** [0.06]	0.005 [0.03]	-0.14*** [0.04]	-0.16*** [0.06]	0.01 [0.03]	-0.15*** [0.03]	-0.11 [0.09]	0.05 [0.04]	0.06 [0.06]
After (Stage 3)	-0.03 [0.03]	0.04*** [0.013]	-0.07*** [0.02]	-0.04 [0.03]	0.01 [0.01]	0.03 [0.03]	-0.04 [0.03]	0.004 [0.01]	0.04 [0.03]	-0.04 [0.03]	0.004 [0.01]	0.04 [0.03]	-0.04 [0.03]	0.004 [0.01]	0.04 [0.03]
Deliberation, \times After				0.09* [0.04]	0.03* [0.02]	-0.12*** [0.03]	0.08* [0.05]	0.04** [0.02]	-0.12*** [0.04]	0.08* [0.05]	0.04** [0.02]	-0.12*** [0.04]	-0.01 [0.05]	0.04 [0.03]	-0.03 [0.04]
Male							0.08 [0.05]	0.02 [0.03]	-0.10*** [0.03]	0.09 [0.05]	0.02 [0.03]	-0.10*** [0.04]	0.09 [0.06]	0.02 [0.03]	-0.11*** [0.04]
Age							0.001 [0.002]	-0.001 [0.001]	-0.002 [0.01]	0.002 [0.002]	-0.002 [0.001]	-0.001 [0.002]	0.003 [0.003]	-0.001 [0.001]	-0.002 [0.002]
Undergrad							0.01 [0.06]	-0.01 [0.03]	0.005 [0.04]	0.006 [0.06]	-0.01 [0.03]	0.006 [0.04]	0.005 [0.06]	-0.01 [0.03]	0.01 [0.04]
British							-0.06 [0.06]	0.04 [0.03]	0.02 [0.04]	0.006 [0.06]	0.004 [0.02]	-0.01 [0.04]	0.005 [0.07]	-0.0004 [0.04]	-0.005 [0.04]
Student										0.14* [0.07]	-0.06* [0.04]	-0.07 [0.05]	-0.10 [0.13]	-0.08 [0.07]	0.18** [0.08]
Student \times Deliberation													0.16** [0.06]	-0.01 [0.03]	-0.15*** [0.05]
Student \times Deliberation \times After													0.53*** [0.15]	0.15** [0.07]	0.32*** [0.11]
Constant	0.66*** [0.04]	0.10*** [0.02]	0.23*** [0.03]	0.70*** [0.05]	0.12*** [0.03]	0.18*** [0.04]	0.69*** [0.11]	0.12** [0.05]	0.19** [0.07]	0.58*** [0.15]	0.19*** [0.08]	0.27*** [0.09]	0.53*** [0.15]	0.15** [0.07]	0.32*** [0.11]
Observations	320	320	320	320	320	320	320	320	320	320	320	320	320	320	320
Adj. R^2	0.01	0.01	0.02	0.01	0.01	0.02	0.03	0.02	0.07	0.05	0.04	0.07	0.06	0.04	0.095

Note: Standard errors clustered at individual level; 'After' = after deliberation. *, **, *** means significant at 10%, 5% and 1% levels respectively. The omitted group (baseline) is the non-deliberation group.

We begin this exercise with a regression (Reg (1)) with no control variables. We find that the coefficient for “Deliberation” (treatment variable) is not significant for any of the preference parameters (θ_i). However, the coefficient for “After Stage 3” (after deliberation) is significant for θ_2 - preference for joint payoff (0.037, p-value = 0.004) and for θ_3 - preference for variance in payoff (-0.07, p-value = 0.005). Since these coefficients pick up the effects in both deliberation and non-deliberation games, we then add, in the next columns (Reg (2)), an interaction term between “Deliberation” and “After” Stage 3 into the regression. The coefficient for the interaction term is significant for θ_3 (-0.12, p-value = 0.03) and θ_1 (0.09, p-value = 0.06) and borderline significant for θ_2 (0.03, p-value = 0.09). In other words, group deliberation significantly reduces the preferences for variance in payoff (θ_3), but increases a concern for own payoff (θ_1) while it has little impact on the preference for joint payoff. Regression (3) adds a set of covariates which include age, gender, undergrad, and being British. The only variable which is significant is male; male subjects are, on average, less concerned about variance in the payoff compared to female subjects. The treatment variable (“Deliberation”) is now significant and negative for preferences for own payoff (θ_1) and preference for variance in the payoff (θ_3), suggesting that subjects in the treatment group initially care less about own payoff and variance in the payoff compared to the control group. However, our main interest is in the interaction term between “Deliberation” and “After Stage 3” which represents the impact of being treated (having the opportunity to discuss the choices with the group), we find that group deliberation increases both concerns for own payoff (θ_1) and joint payoff (θ_2), but makes the subjects care less about variance in the payoff (θ_3).

In regression (4), we include the subject pool dummy (student = 1, non-student = 0) and in regression (5), we include an interaction term between students and the treatment (deliberation) and a three-way interaction term between students, deliberation and the after deliberation (stage 3). The objective here is to examine whether the treatment influences students and non-students differently. Looking at regression (4), all the treatment effects remain robust and students overall appear to care more about their own payoff compared to non-student subjects and are less concerned about joint payoff, although the coefficients are only marginally significant. Once we add the interaction terms between student and treatment (Deliberation) and the three-way interaction term between Student, Treatment, and After Stage 3, the main treatment variables (Deliberation and Deliberation \times After Stage 3) are no longer significant. Instead, the three-way interaction term is significant for preferences for own payoff (θ_1) and variance in the payoff (θ_3) and thus, our results suggest that students are more affected by group deliberation than non-students. Specifically, students become more selfish and less concerned about variance in the payoff.

We also ran additional specifications which included other interaction term between student and gender

and a three-way interaction term between student, gender and treatment (deliberation) and the results remain robust (available upon request). Overall, we find that our group deliberation treatment affects the preferences for own payoff (θ_1) and variance in the payoff (θ_3), but not on the preference for joint payoff (θ_2). When we examine the subject-pool specific effect (students vs. non-students), we find that these effect are driven by the student subjects.

5.2 Within Group Preference Heterogeneity

We now move on to our main analysis which is to investigate whether there is any evidence of changes in within-group preference homogeneity. In order to do this, we make use of the fact that our empirical approach allows us to locate each individual within a well-defined parameter space and thus we are able to measure the distance between their preferences and those of others. This provides us with our measure of inter-personal preference heterogeneity.

Figure 5: Group 6 (five members)

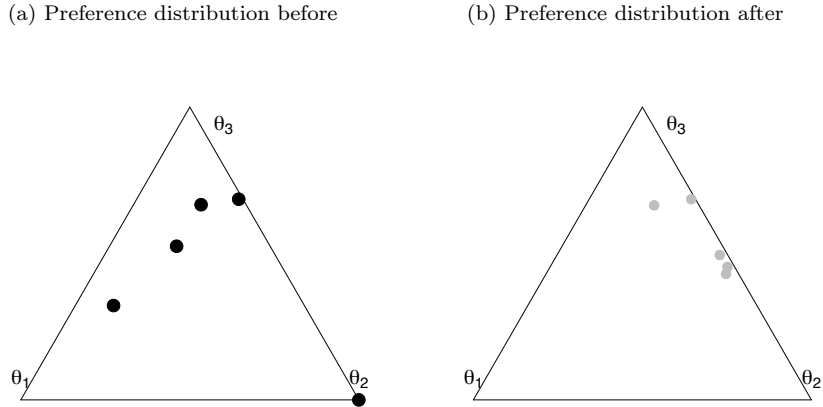
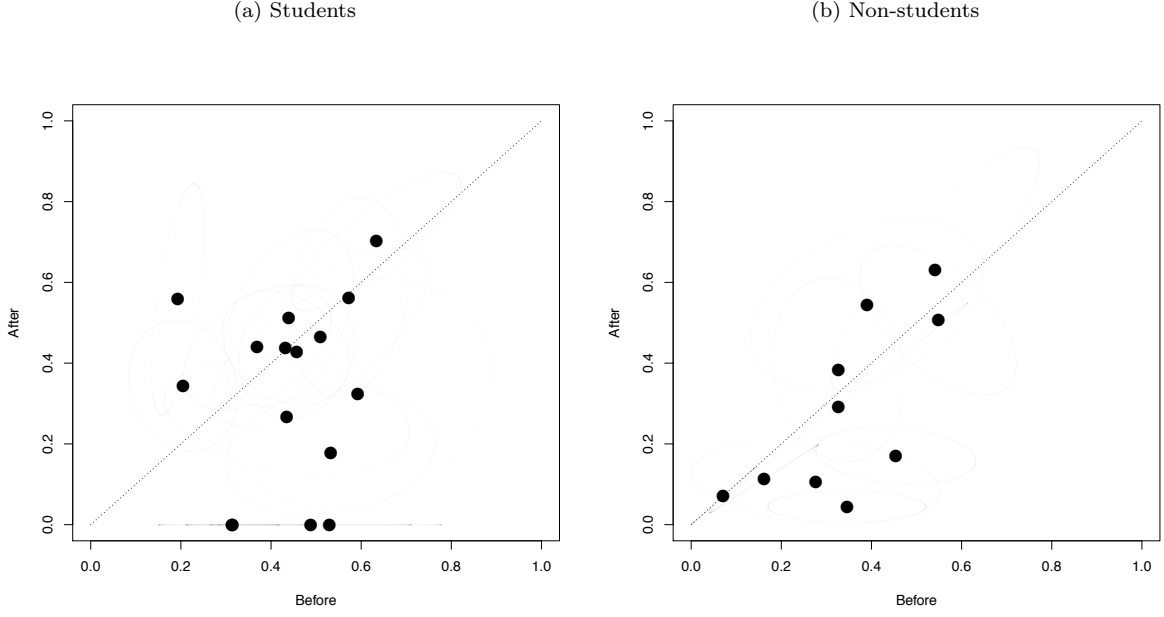


Figure 5 illustrates this idea for group 6 of the non-student subject pool. There were five members of this group. In figure 5a each black dot indicates a subject's preference parameters prior to the group interaction phase; the grey scatter in 5b shows them afterwards. Before the group interaction, the preferences of the group were evidently spread around the parameter space with one individual being a clear outlier with preferences that were entirely centered on the joint payoff (bottom right). After the group interaction, the preferences can be seen to be more tightly bunched within the parameter space (afterwards this group generally moved away from egotism). We take this to indicate a reduction in preference heterogeneity within the group.

Figure 6: Preference Heterogeneity, before and after, by group



In order to summarize this change in preference heterogeneity, we used the average (Euclidean) distance between the preferences of group members before and after social interactions. The mean distance is bounded in our parameter space by the unit interval. For the group illustrated in Figure 5(a),(b) for example, we measure preference heterogeneity at 0.454 before the interaction and 0.171 afterwards. This fall reflects the bunching we see in the figure.

We calculate these group preference heterogeneity measures for both of our treatment groups (students and non-students) before and after the interactions phase and compare them in Figure 6 (with their 95% confidence ellipses). Figure 6a shows the results for the students²³ and 6b for the non-students. Both for the students and the non-students, most of the groups lie either below or close to the 45-degree line indicating that preference heterogeneity tends either to decline or remain roughly the same after the social interactions. In other words, the preferences of people who interacted in the same group became more similar after social interactions i.e. there is evidence of ‘preferences conformity’.

Table 6: Changes in preference heterogeneity

	All	Students	Non-students
Mean change in heterogeneity	-0.0922	-0.1124	-0.0575
std error	0.0411	0.0536	0.0702
p-value	0.0341	0.0422	0.411

Notes: clustered standard errors.

Table 6 summarizes the overall effects on group heterogeneity. Overall there is a significant increase in homogeneity of preferences and the mean distance between individuals within each group falls by 0.0922. However, this is mainly being driven by the student subjects (the non-students also see increased homogeneity but this is more modest and is not statistically significantly different from zero).

²³Note that there appear to be only 15 points plotted in Figure 10(a) even though there were 16 groups. This is because two of the groups lie almost exactly on top of each other (at (0.31,0)).

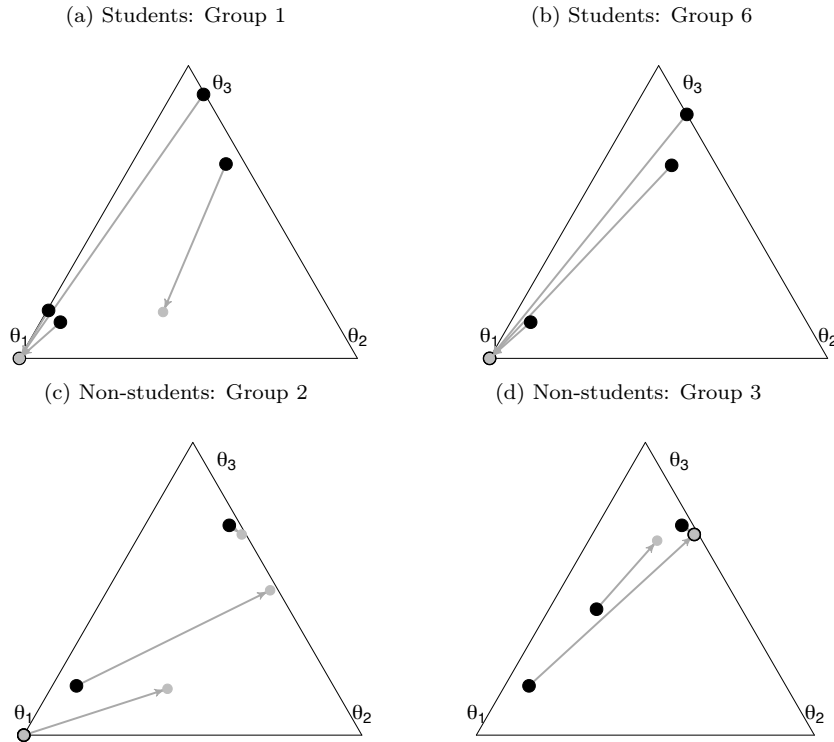
5.3 Group Preference Dynamics

What or who drives this increased conformity, particularly amongst the student group, is an open question as is what type of preference people ‘conform’ to. In order to shed some light on this, we next examine the dynamics of preference changes within groups.

To illustrate our approach, consider Figure 7. This shows four groups and the preferences for the members of each group before and after social interactions/group deliberation are illustrated. A black dot indicates a subject’s preference parameters prior to the interactions phase; a grey dot shows them afterwards. The arrows connect the before-after parameters for each individual. If the subject’s preferences didn’t change, the grey dot (the ‘after’ preference) is super-imposed on the slightly larger black dot (the ‘before’ preferences). Figure 7 shows, firstly, whose preferences changed and whose did not, and, secondly, how preferences change.

If we consider, for example, student group 1, the preference parameters of four of the group members changed. One subject’s preference did not change: (s)he was located in the bottom left corner and thus had strongly egoistic preferences ($\theta_1 = 1$). Three of the other subjects (two of whom were already displaying a large value for θ_1) moved towards this individual and joined her/him in the bottom left corner with fully egoistic preferences. The fourth group member also moved towards a more strongly egoistic preference type afterwards but did not go as far as the others. Student group 6 shows a similar pattern but more strongly - *all* of the group members move towards an individual located in the strongly egoistic corner. For the two non-student groups illustrated we also see some interesting dynamics. In the case of Group 2 we see an individual who displays no egotism at all who remains in the same place after the deliberation phase, whilst others move towards him/her. Group 3 shows two individuals (again both entirely lacking in egotism) who do not change their preferences whilst others in the group move towards them.

Figure 7: Preference dynamics - Examples



These examples show three interesting features of the distribution of preferences and of dynamics.

- Firstly, initial preferences within groups were evidently heterogeneous (from the scatter of initial preferences across the parameter space).
- Secondly, the changes in preferences are also heterogeneous (some individuals change, others do not).
- Thirdly, subjects who had extreme preferences seem to stick to their guns whilst other subjects were observed to adopt the same preferences after the interactions phase.

We have plotted in Appendix C the corresponding diagrams showing the preference dynamics for *every* group. These show the preferences of members before and after the exposure to social interactions/group deliberation. If we look across the rest of the groups of student subjects, this last feature (someone with rather extreme preferences attracting others towards them) is clearly noticeable in several groups. This is tantalizing evidence of what might be termed *the attractive qualities of extremes*. Appendix C discusses the individual group dynamics in detail.

To examine this observation further, we define a subject to be an “Attractor” if that individual’s preferences do not change and everyone else in the group moves closer to them. If there are two such people then the one closest to the others in the group after the interaction is defined as the Attractor. Groups can have two Attractors if there are two people with the same unchanging preferences who both, therefore, fulfill the minimum distance desiderata.

The majority of the groups (60%) possess Attractors (the proportions are almost exactly the same in the student and non-student subject groups). Table 7 uses a regression to describe the preferences of Attractors. It is notable that, whilst the Attractors in the student group are on average quite extreme in their preferences with regard to their insensitivity to inequality (θ_3) and tend towards egoism, the older, non-student subjects who act as Attractors are not so extreme and exhibit more balanced preferences.

Table 7: Preference Parameters Attractors and Followers

Dependent variables:	Own payoff (θ_1)	Joint payoff (θ_2)	Inequality in payoffs (θ_3)
Attractor	0.0924 [0.2319]	-0.0765 [0.0859]	-0.0159 [0.1543]
Students	-0.0075 [0.1204]	-0.0731 [0.0642]	0.0806 [0.0718]
Attractor×Students	0.3199* [0.2423]	0.0037 [0.0969]	-0.3236** [0.1610]
Constant	0.5373*** [0.1096]	0.1898*** [0.0610]	0.2729*** [0.0559]
N	127	127	127
$F(3, 123)$	4.875	2.243	4.8696
Adj R-square	0.10626	0.05187	0.10667

Notes: N=127; standard errors clustered at the group level.

*, **, *** indicate significant at 10%, 5% and 1% levels respectively.

Table 8 investigates the relationship between the presence of extreme preferences within a group and the effects on preference heterogeneity. In particular we are interested in whether the observation made above and in Appendix C that group interactions are associated with a convergence in preferences towards those of relatively extreme individuals (especially among the younger student group) is a robust feature of the data.

Table 8 describes the relationship between the presence of an Attractor with extreme preferences on the associated change in the group’s preference heterogeneity. We define an Extremist to be an Attractor for whom one of their preference parameters gives a weight to a particular aspect of the division of resources greater or

Table 8: The Presence of Extremists and Group Heterogeneity

Dependent variable:	Group Heterogeneity
Extremist Present	0.1082 [0.1030]
Students	0.1620 [0.0847]
Extremist \times Students	-0.3691*** [0.1214]
Constant	-0.1716** [0.0847]
N	30
$F(3, 26)$	3.691
Adj R-square	0.2987

Notes: N=30; standard errors are at the group level.

*, **, *** mean significant at 10%, 5% and 1% levels respectively.

equal to 0.95.²⁴ In these data defining extremists according to this criteria means that all of the extremists turn out to be egoists rather than, for example, extreme egalitarians. The dependent variable is the change in the group preference heterogeneity (measured as above). The regression allows for heterogeneous effects according to whether the group consisted of the student subjects or the non-students. The homogenizing effect of the presence of someone with extreme preferences on the group is clearly a phenomena which is present among the students - there is a converging effect on preferences which is both significant in magnitude and also robust to sampling variation - but not operating among the non-students.

One possibility which might explain some of the regularities described above is that there could be a natural tendency toward egoism when choices are repeated even without deliberation. If so then this would give rise to people who appear to look like egoistic "attractors"—they remain egoistic and others change their behavior in this direction. That is, with any absorbing state in a dynamic process, the first people who arrive will look like "attractors" even if they have no impact on why others move there.²⁵ To investigate this, Appendix C (Figure C.3) also shows the preference dynamics for the non-deliberation group and a table (Table C.1) describing the changes in preferences which occur for these individuals. Overall there is no evidence of significant changes in any particular direction for these subjects: on average the largest preference shift (albeit statistically insignificant) is in egotism but, rather than towards greater self-interest, it is away.

6 Discussion and Conclusion

In this paper, we study the extent to which social interactions by means of face-to-face communication/deliberation affect an individual's other-regarding preferences. Specifically, the question of interest is whether and how people's preferences change *once they are away from the group* and have to make their own decisions in private without the presence of the group. Our main findings are as follows.

In the non-deliberation group where there is no social interaction, when asked to make the same choices three times, subjects exhibit reasonably consistent and stable choices and hence choice-revealed preferences. In our treatment/deliberation groups, we find evidence that the student subjects show more behavioral and preferential change after the interactions than the control/non-deliberation group and non-student group.

We find that both initial (pre-interactions) revealed preferences and the preference changes that we observe after interactions are *heterogenous*. Some people become more selfish, whilst others care more about total payoff or equality in the payoffs (and some people, on the other hand, do not change their preferences at all) and thus, at aggregate level, heterogeneity of the effects may to cancel each other out. Indeed, analysis at individual level (and on a group-by-group basis) reveals that in most groups, preferences of the group members become

²⁴The results reported are robust to less stringent definitions of extremism.

²⁵We are very grateful to an anonymous referee for making this observation and for suggesting the following checks.

more homogenous after social interactions even after they are away from the group (and thus, they were not subject to monitoring or being observed by the other group members). More importantly, we find that students are more likely to change their preferences compared to non-students and their preferences are influenced by individuals who have *extreme* preferences (in most cases, extremely selfish or egoistic) that do not change after interactions. We refer to these individuals as ‘Attractors’ since they appear to be pulling others’ preferences towards their own. We find that attractors tend to be male, at least amongst student subjects. Amongst non-student subjects, the influence of the Attractors is less pronounced. This is partly due to the fact that the non-student Attractors do not share the same preferences (even when they are in the same group) and there are more than one Attractors in most groups. In groups with no Attractor, social interactions also change preferences, but the changes go in different directions. As a result, we do not observe preference convergence in these groups.

We find it interesting that most subjects, particularly the students, converge towards extremely egoistic/selfish preferences after the group interactions. A number of potential mechanisms could be driving this result. For example, public deliberation about what action to take could influence individuals within the group to do ‘the right thing’ (whatever that may be) rather than making choices that reflect their true preferences. Social interactions also allow the group members to learn information about the distribution of others’ preferences (as posited by the informational cascade literature) and once they have learned that others have self-interested or egoistic preference, they may feel less obliged to be nice to the receiver (the ‘moral cost’ of being selfish is reduced). It is possible that a group norm emerges during the interactions and in the case where selfish behavior is established as a norm, the group members may feel obliged to conform to this norm for fear of being stigmatized (Young, 2014) or face with a threat of sanctions (Fehr and Gächter, 2000; Fischbacher, Gächter and Fehr, 2001) even after they have left the group. Moreover, since the task is novel and the decisions are made in a somewhat unnatural context (playing dictator games), where individual preferences are themselves weak and/or not well-established, subjects may try to establish an agreement on what would be a ‘norm’ in this context during the deliberation phase. However, it is worth pointing out that the selfish norm was not established among non-student subjects. The fact that they work in the same organization and are more likely to meet each other again after the experiment might play a role in determining the agreed norm within the group. Finally, it is possible that since the Attractors have strong and extreme preferences that are unchanging, they are able to convince or persuade others to follow them. Furthermore, since behaving selfishly also maximizes one’s own payoff, it may not take much effort to convince the other members to change the behavior compared to the effort required for inequality aversion and social efficiency preferences which may involve decreasing one’s own payoff. In our experiment setting, we cannot identify which of these mechanisms drive the group decisions and thus, our analysis focuses on individual decisions pre- and post-deliberation.

We neither recorded nor monitored what subjects said during this stage because we wanted to keep the interactions as natural and as free-flow as possible in order to avoid social desirability bias and experimenter’s demand effect. Subjects may produce answers that they think are correct, socially acceptable, or what the experimenter wants, if they think they are being “watched” or that what they say is recorded and judged. We also did not impose any decision rules and let the subjects arrive at the group decisions organically. The trade-off is that we did not know the precise mechanism that was driving the group decisions e.g. whether one person dictated the group or whether majority rule applies.

This procedure, however, serves the main objective of the current paper, which is to examine whether and how group interactions/deliberation affects individual preferences and not what happens during the group decisions stage which has already been examined by a large number of studies (Alpizar et al., 2008; Bardsley and Sausgruber, 2005; Cason and Mui, 1998; Charness and Sutter, 2012; Cooper and Kagel, 2005; Frey and Meier, 2004; Levitt and List, 2007; Luhan et al., 2009; Shang and Croson, 2009; Vesterlund, 2003; 2006).

One could argue that the observed strong influence of selfish Attractors among students may be due to the larger proportion of economics students in this subject pool. We were very careful in our recruitment process

to avoid economics and psychology students who are more experienced with game theory and experimental methods. Unfortunately, we only collected data on the levels of education (undergraduate, masters, PhD) and not on the fields of study. Thus, we are not able to analyse the specific impact of economics students. However, non-student subjects were mainly professional auditors and accountants, who were likely to have some economics training or have an economics degree. Thus it is possible that there were more economists amongst non-student subjects than the students who were randomly recruited from a university-wide subject pool.²⁶

It would be interesting for future research to study the mechanisms that cause the changes in preferences during group deliberations by recording the conversations that take place within the group and run additional treatments to gain a better understanding of the group dynamics, such as varying decision-making rules and punishments and looking into the social interaction processes i.e. how the Attractors manage to change others' preferences. For example, whether the Attractors are more likely to be persuasive than other group members or assumes a 'leader' position in making the group decision. It would also be interesting to investigate how different group sizes affect social interactions and whether the role of the extremist Attractors may change within larger groups. Finally, it would be useful to examine whether a policy intervention that changes the preferences and behavior of the Attractors (for example, from extremely selfish to more altruistic) can have a spillover effect for the rest of the group. Policies that target these key individuals may be more effective than those that attempt to change the behavior of the group as a whole.

²⁶We are grateful to an anonymous referee for pointing this out.

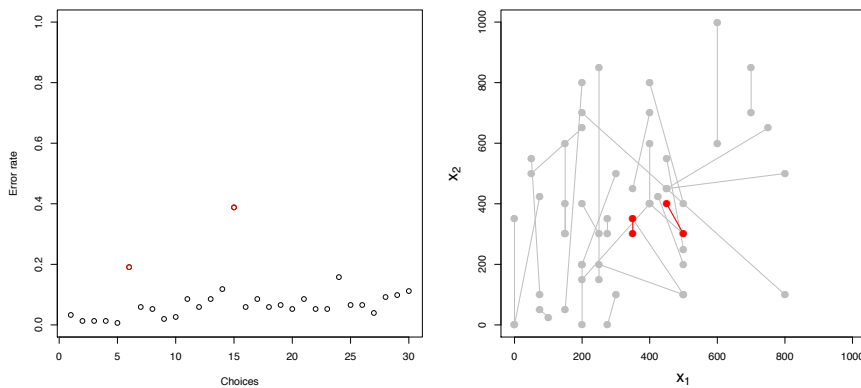
Appendix A - Non-rationalizable Choices

Rather than appending an error term to our model and fitting it by minimizing the sum of squared errors, in our empirical work we choose the best-fitting model according to our alternative loss function which is the Selten Index of predictive success. Part of this loss function is the pass rate or the largest number of an individual subject's choices that the model can jointly rationalize (sometimes called the Houtman-Maks index). We then retain only those observations.

The following Figures A.1 to A.3 provide information on the observations which we discard in our approach, by each of the three subject groups. In each case panel (a) illustrates the proportion of times each choice was removed from the data on the grounds of non-rationalizability ("error rate") where the x-axis is each choice (numbered as in Table 1), and panel (b) illustrates which pairs of choices were deleted from our data most frequently.

Figure A.1 shows the student subjects. Overall, there seems to be no obvious systematic pattern of non-rationalizability and, were it not for two outliers, one might conclude that non-rationalizability was occurring at random with respect to the choice problem, and hence that when non-rationalizable observations are removed they become missing at random. However, we do see two clear outliers. These are the choices highlighted in panel (b): choices 6 $\{(350, 350) \text{ or } (350, 300)\}$ and 15 $\{(450, 400) \text{ or } (500, 300)\}$. A common feature of these choices is that in one of the options the payoff difference between the sender and the receiver is very small i.e. only 50 EMUs (or 50 pence). For example, for choice 6, the sender has to choose between an equal payoff distribution option and an alternative which yields an almost equal distribution. Therefore, inequity averse individuals may be indifferent between the two options and thus, make their choices at random, resulting in non-rationalizability. In addition, the sender also receives the same payoff regardless of the option that they choose and thus, senders who care only about their own payoff may also be indifferent between these two options. A similar argument can also be applied to choice 15 which has a higher error rate (40%) where the difference between the sender's and the receiver's payoffs is also small in both options and there is not much difference in the sender's own payoff either.

Figure A.1: Non-rationalizable Observations, Students
(a) Error Rates (b) Choices



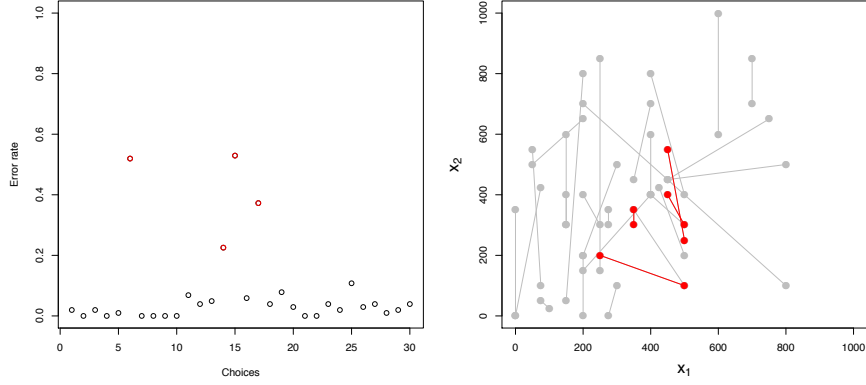
For non-student subjects, there are four choices which are non-rationalizable as shown in Figure A.2. Two of these are choices 6 and 15 which may be due to the same reason as outlined above. The other two choices are choices 14 $\{(500, 250) \text{ or } (450, 550)\}$ and 17 $\{(500, 100) \text{ or } (250, 200)\}$. For choice 14, the payoff to the sender is, again, very similar between the two options (500 vs 450), although there is a large difference in the payoff that the receiver could potentially get (250 vs. 550). For choice 17, it is the opposite i.e. the difference in the payoff that the senders could allocate to themselves between the two options is large (500 vs. 250), whilst the difference in the payoff that the receiver could potentially get is smaller (100 vs. 200). In addition, the payoff difference

between the sender and the receiver is also much larger in one option ($500 - 100 = 400$) compared to the other ($250 - 200 = 50$). Since the differences are more distinctive (compared to the other three non-rationalizable choices), it is more difficult to speculate why choice 17 is not rationalizable for non-student subjects.

Figure A.2: Non-rationalizable Observations, Non-students

(a) Error Rates

(b) Choices

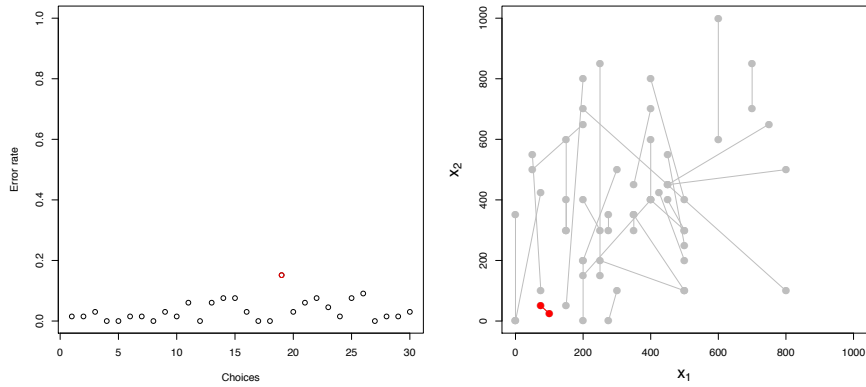


For the control/non-deliberation group, it is not entirely clear that there indeed was a particular choice which tended to be non-rationalizable. The only choice that failed to be rationalized more often than others was choice 19 $\{(75, 50) \text{ or } (100, 25)\}$. In this case, the difference in the payoff that the senders could allocate to themselves is 25 EMUs which is also the same as the difference in the potential payoff that the receiver could get. The total endowment in both options is the same. Since the error rate for this choice was low (less than 20%), it is possible that the error was caused by subjects simply making a mistake.

Figure A.3: Non-rationalizable Observations, Non-deliberation Group

(a) Error Rates

(b) Choices



Appendix B - Preference Distributions

We also examine preference distributions of both deliberation groups and the non-deliberation group in stages 1 and 3.

Table B.1: Descriptive Statistics, Preference Parameters: Students

		Mean	Std. Dev	10th percentile	Median	90th Percentile
<i>Before</i>	θ_1	0.6058	0.3663	0.0221	0.7135	1
	θ_2	0.1032	0.1647	0	0.0536	0.2793
	θ_3	0.2910	0.3278	0	0.1237	0.8333
<i>After</i>	θ_1	0.7210	0.3591	0.1337	1	1
	θ_2	0.1412	0.2199	0	0	0.5068
	θ_3	0.1379	0.2550	0	0	0.4898

Table B.1 shows preference distributions for *student* subjects. There was an increase in the average of θ_1 from 0.6958 to 0.7210. There was a smaller increase in the average preference for the joint payoff (0.1032 to 0.1412) and a marked decline in inequality aversion amongst this group (θ_3 fell from 0.291 on average to 0.1329). A statistical test confirms that the changes in the student treatment group's preferences for their own payoff and their inequality aversion were both statistically significant (Wilcoxon sign-rank test: $z = 2.51$, $p = 0.01$ and $z = 4.11$, $p = 0.00$ respectively).

Figure B.1: Changes in Preferences, Student Subjects

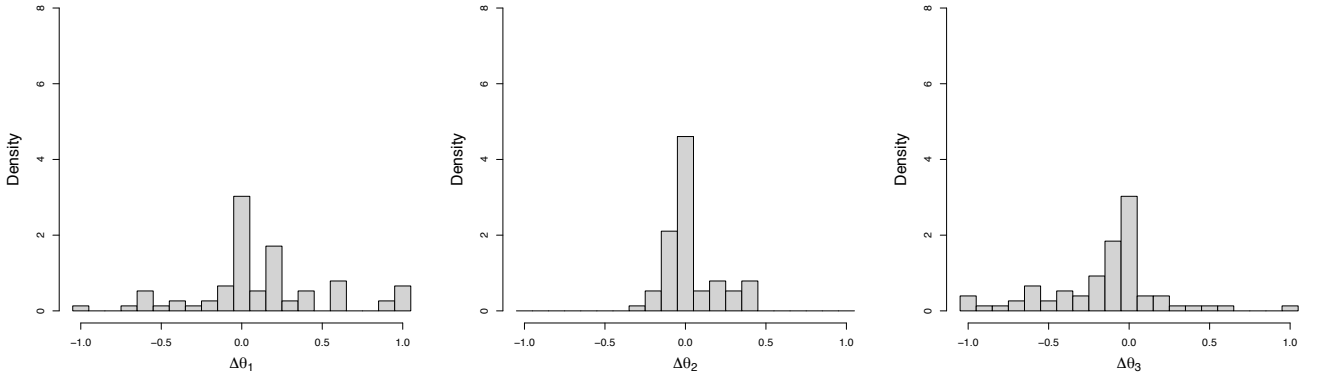


Figure B.1 shows the distribution of changes in preference parameters (scaled to represent a density, as are the other figures in this section). The mass point at $\Delta\theta_1 \approx 0$ shows that many subjects' preference for their own payoff did not in fact change at all. There is significant heterogeneity in the distribution of changes with some subjects having a change in this parameter $\Delta\theta_1 \approx -1$ indicating that while they started off very selfish ($\theta_1 \approx 1$) they became much less so ($\theta_1 \approx 0$), whilst others did the opposite. As can be seen from the histogram of $\Delta\theta_1$, there is somewhat greater weight of data on the positive side, indicating that, on average, the change in the preference for own-payoff for the students to give it more weight after the social interactions/group deliberation. If we look at the histograms relating to preference for the joint payoff (θ_2) and the inequality in payoffs (θ_3) we see corresponding changes. In both of these distributions there is clear heterogeneity, but also mass points around 0 which also become more pronounced as preferences change towards giving more weight to own-payoff.

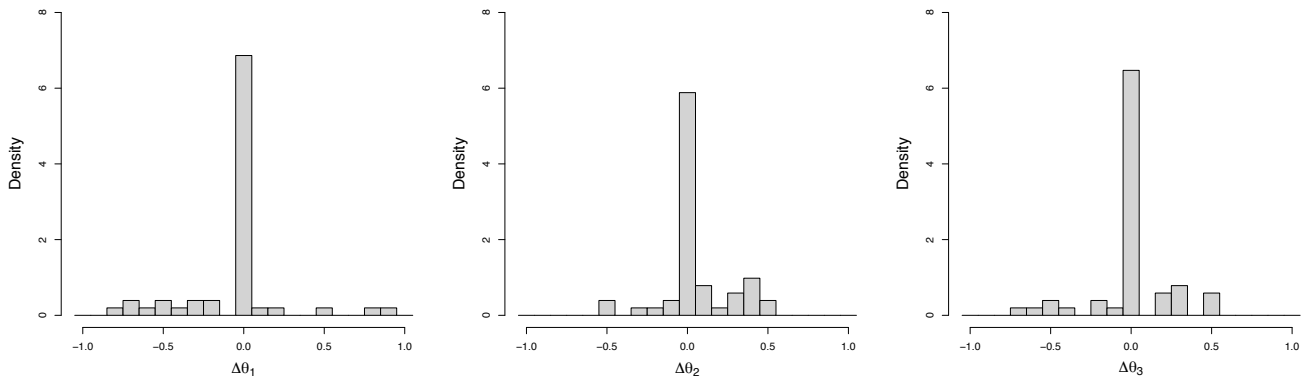
Next, we look at *non-student* subjects' preferences (Table B.2, Figure B.2). Non-students appear to be less selfish compared to student subjects to begin with and largely remain so. The only significant change in average preference parameter values is in the average θ_2 from 0.1658 to 0.2186 (Wilcoxon sign-rank test: $z = 2.07$, $p = 0.04$) indicating something of an increase in preference for efficiency (total payoff).

Table B.2: Descriptive Statistics, Preference Parameters: Non-students

		Mean	Std. Dev	10th percentile	Median	90th Percentile
<i>Before</i>	θ_1	0.5663	0.4357	0.0126	0.7600	1
	θ_2	0.1658	0.2260	0	0.0717	0.3020
	θ_3	0.2679	0.2980	0	0.1683	0.6854
<i>After</i>	θ_1	0.5118	0.4664	0.0126	0.6481	1
	θ_2	0.2186	0.2568	0	0.0758	0.5245
	θ_3	0.2696	0.2917	0	0.1583	0.6854

The distributions of changes in preference parameters for these subjects are show in Figure 9. Similar to student subjects, the mass point at $\Delta\theta_1 \approx 0$ suggests that many subjects' preference for own payoff did not change. When we examine the distribution of changes in the preference for own-payoff ($\Delta\theta_1$), there is greater weight of data on the negative side, suggesting that non-students became less selfish after the group deliberation. Whilst the preference for total payoff for most subjects did not change (a spike at $\Delta\theta_2 \approx 0$), there is more mass on the positive which indicates that non-students became more concerned about the joint payoff after interacting with each other. Finally, their preference for equality in the payoff distribution (θ_3) did not change: the pronounced mass point at $\Delta\theta_3 \approx 0$ in the third histogram. However, there is heterogeneity in the distribution of the changes as some subjects became less concerned about inequality $\Delta\theta_3 < 0$, whilst others changed in the opposite direction. A comparison of the student and non-student distributions of changes in preferences shows that the students change more - there is less density at zero and more elsewhere.

Figure B.2: Preference Distributions for Non-student Subjects



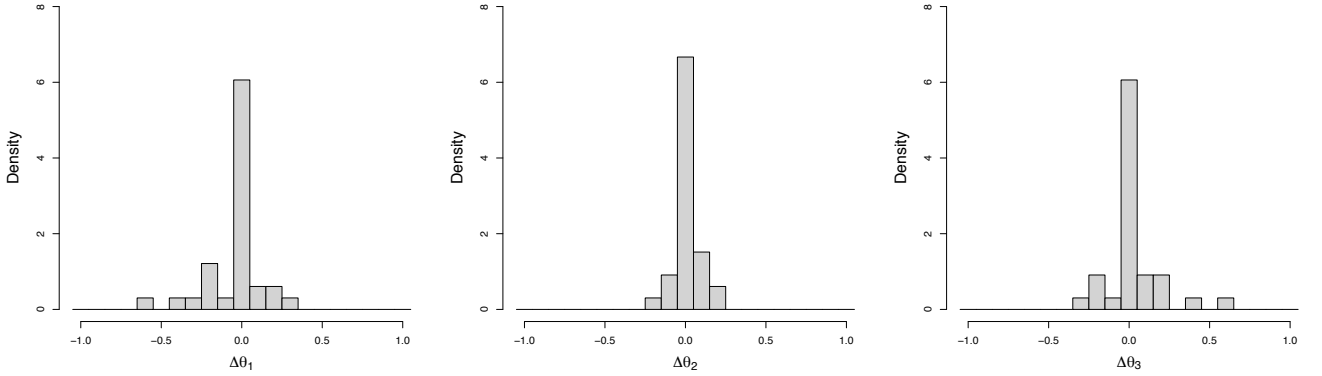
Finally, in the control/non-deliberation group (Table B.3, Figure B.3) we see that these subjects tended to be less generous than the others to begin with and there were no significant change in the average values of these parameters when they repeated their choices.

Table B.3: Descriptive Statistics, Preference Parameters: Non-deliberation group

		Mean	Std. Dev	10th percentile	Median	90th Percentile
<i>Before</i>	θ_1	0.7029	0.2701	0.379	0.8168	1
	θ_2	0.1157	0.1346	0	0.0596	0.2815
	θ_3	0.1814	0.2089	0	0.1237	0.3470
<i>After</i>	θ_1	0.6644	0.2647	0.2306	0.8168	0.8168
	θ_2	0.1248	0.1304	0.0159	0.0596	0.2976
	θ_3	0.2108	0.1937	0.0869	0.1237	0.5202

Since there was no social interactions in this group, we expect to see minimum change in preference parameters, assuming that the subjects' preferences are consistent. Indeed, as shown in Figure B.3, for all three parameters, there is a mass point at $\Delta\theta_{1,2,3} \approx 0$ indicating that subjects in the non-deliberation group did not change their preferences when asked to make the same choices again.

Figure B.3: Preference Distributions for Control/Non-deliberation Group



The general tenor of the changes described seems to be that students change their preferences the most and the non-deliberation group and the non-students the least or not at all. The effects on the means are small as changes in each direction seems to be modest or to cancel each other out. Only for the students does there seem to be a significant “average treatment effect” which has the net result of making them more selfish.

Appendix C - Preference Dynamics

Preference dynamics - students

The preference data for all student groups is provided in Figure C.1. There are 16 groups in total and, as can be seen by studying the figures, 10 groups have at least one Attractor (there were 14 Attractors in total across all groups and most of them were male): groups 1, 5, 8, 11, 15 and 16 have one Attractor; and groups 2, 6, 10, and 12 had two Attractors each. Groups 3, 4, 7, 9, 13 and 14, on the other hand, have no Attractor. The most interesting feature appears to be that most of these Attractors have extremely egoistic preference and they appear to change the preferences of other group members towards this extreme (bottom-left corner).

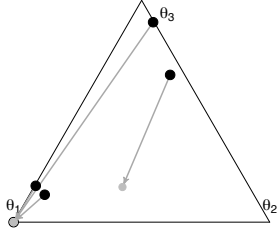
First, we look at groups with one Attractor. For example, group 1 has five members (all male). The Attractor (#119)'s preference is at the bottom-left corner ($\theta_1 = 1, \theta_2 = 0, \theta_3 = 0$) both before and after social interactions. There are four other members in the group: before social interactions, two people (#1 and #15) also place more weight on their own payoff than the total payoff and inequality in payoff distribution (for #1: $\theta_1 = 0.831, \theta_2 = 0.004, \theta_3 = 0.163$; and for #15: $\theta_1 = 0.816, \theta_2 = 0.059, \theta_3 = 0.123$). The other two (#7 and #56) care more about inequality, followed by the joint payoff and then own payoff (for #7: $\theta_1 = 0.004, \theta_2 = 0.093, \theta_3 = 0.901$; and for #56: $\theta_1 = 0.056, \theta_2 = 0.279, \theta_3 = 0.664$).

However, after social interactions, all group members become extremely selfish ($\theta_1 = 1$), except for #56, but his preferences also move towards the bottom-left corner ($\theta_1 = 0.496, \theta_2 = 0.346, \theta_3 = 0.158$ compared to before interactions: $\theta_1 = 0.057, \theta_2 = 0.279, \theta_3 = 0.664$). Thus, the preferences of the whole group converge towards the Attractor's egoistic preference. The same dynamic is also observed in groups 5 and 11 where the Attractors have extremely egoistic preference ($\theta_1 = 1$) and after social interactions the preferences of the rest of the group converge perfectly towards the bottom-left corner. For groups 8, 15 and 16 we also see some movements towards more egoistic preference, but not for everyone in the group.

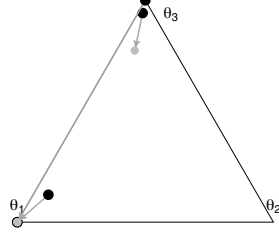
For groups with two Attractors (2, 6, 10, and 12), in all but one group (group 10) the Attractors also have extremely egoistic preference. In groups 2, 6 and 12 the preferences of all group members converge perfectly towards the bottom-left corner. In group 10, however, one of the Attractors has a strong preference for total payoff. Since this group only consists of four members, after social interactions one of the other member moves towards the egoistic Attractor, whilst the other moves towards the Attractor who cares about total payoff.

It is interesting to note that in groups where the preferences converge perfectly (groups 5, 6, 11, and 12) as well as the groups that converge almost perfectly (only one person does not converge perfectly i.e. groups 1 and 2) at least one other group members, who is not an Attractor, already places a stronger weight on own payoff (θ_1). Hence, it may not require much effort to move towards extremely egoistic preference. Furthermore, in groups that the members' preferences perfectly converge, either the majority or all group members are *male*. In the other six groups that have no Attractor (groups 3, 4, 7, 9, 13 and 14), the gender composition is more mixed and in most cases there are more women in the group.

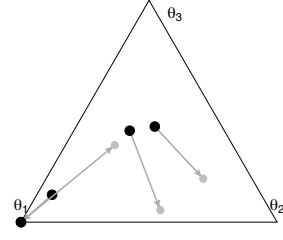
Figure C.1: Within-group Changes in Preferences (Students)



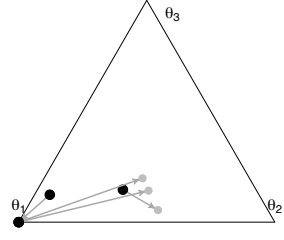
(a) Group 1



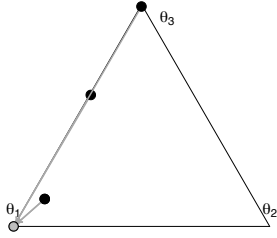
(b) Group 2



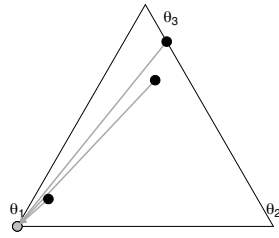
(c) Group 3



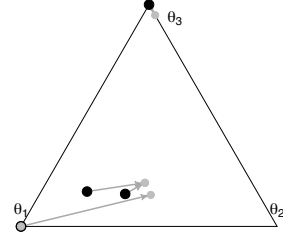
(d) Group 4



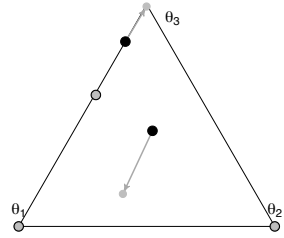
(e) Group 5



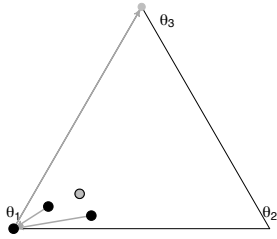
(f) Group 6



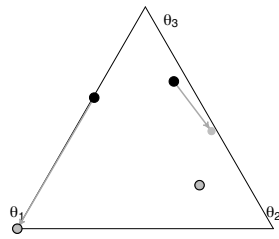
(g) Group 7



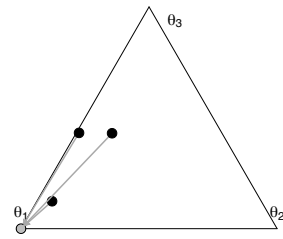
(h) Group 8



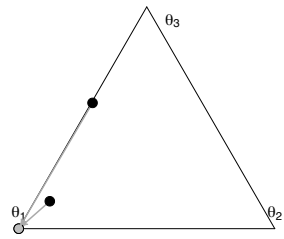
(i) Group 9



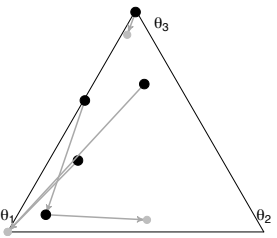
(j) Group 10



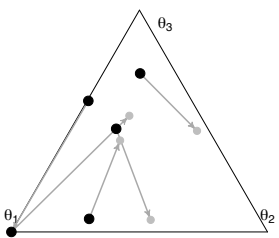
(k) Group 11



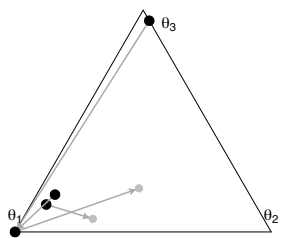
(l) Group 12



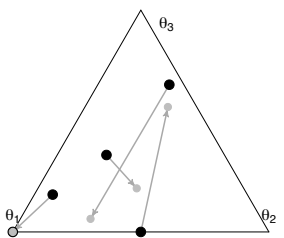
(m) Group 12



(n) Group 13



(o) Group 15

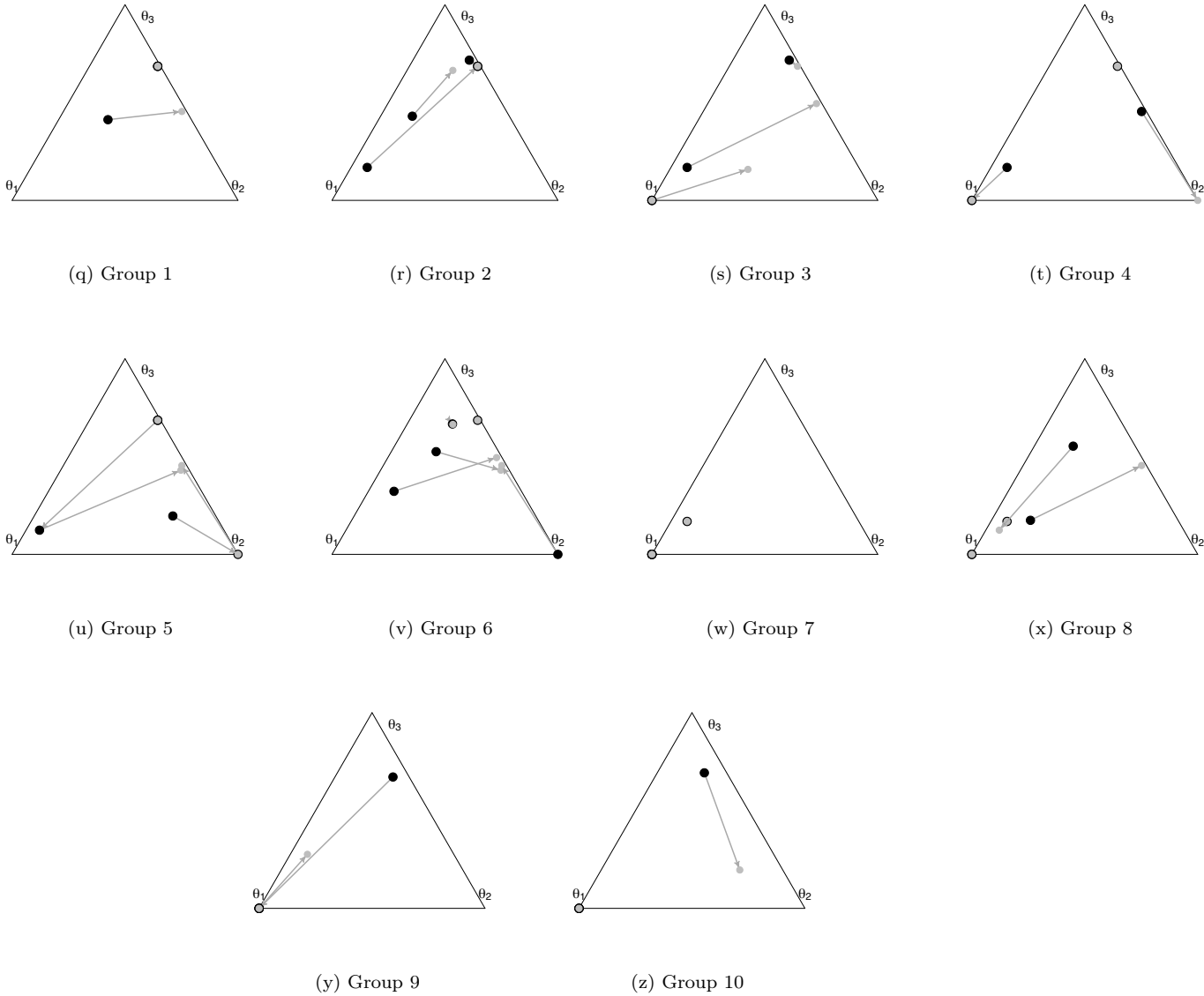


(p) Group 16

Preference dynamics - non-students

There are 10 groups in total in the non-student subject pool (Figure C.2) and 16 Attractors spread across 6 groups (groups 1, 2, 4, 6, 9 and 10). Only group 6 has one Attractor and the rest have at least two. In contrast to student subjects, non-student Attractors do not all share the same preference. Ten out of 16 Attractors have a strong egoistic preference, but the others have a strong preference for the equality in the payoffs. There is also more heterogeneity in terms of the gender composition (10 Attractors are female and 6 are male).

Figure C.2: Within-group Changes in Preferences (Non-students)



For preference dynamics, first we look at group 6 which has only one Attractor who is a male and has a strong preference for equality in payoffs ($\theta_1 = 0.013, \theta_2 = 0.302, \theta_3 = 0.685$). As shown in Figure C.2, there are some movements towards the Attractor but the preferences of group members do not converge. Subjects #112, #113 and #114 all move closer to the Attractor, but #110 does not move much at all (this is shown by a small arrow on top of the grey/black dot). In other groups, which have more than one Attractor, preferences of the

group members move much closer towards the Attractors. For example, in group 2 where there are two Attractors (one male and one female, both care the most about equality in the payoffs), after social interactions all other members converge to the Attractors' preferences ($\theta_1 = 0.013, \theta_2 = 0.302, \theta_3 = 0.685$), except for #129 whose preferences also move closer towards the Attractors' but does not fully converge. Therefore, unlike student subjects, we do not see an absolute convergence of preferences in non-student groups. This is true even in groups where the Attractors have extremely egoistic preference (groups 4, 9, and 10). There is always at least one person in the group who either does not move or moves in the opposite direction away from the Attractor's preferences.

For the groups with no Attractors (groups 3, 5, 7, and 8), preferences appeared to be moving in all different directions. Group 7 is an interesting case where most group members have extremely egoistic preference (5 out of 6 people have $\theta_1 = 1$), except one subject (#140), but he/she also places a strong weight on own payoff ($\theta_1 = 0.76, \theta_2 = 0.072, \theta_3 = 0.168$). In this group, no one moves after social interactions but since preferences are already similar prior to deliberation, social interactions may just reinforce their original preferences. Visual inspection of the groups' parameters spaces shows that Attractors may have a variety of preferences, but also that there is some evidence that extreme, egoistic preferences are quite common.

Preference dynamics - non-deliberation group

The figure shows the preference dynamics for the non-deliberation group. Note that there appear to be fewer than 33 subjects. This is because some of them made the same choices, their revealed preferences are the same and therefore so are their locations in the parameters space.

Figure C.3: Changes in Preferences , non-deliberation/control group

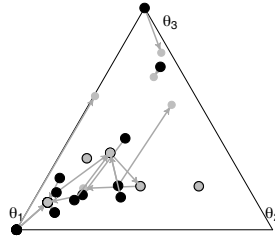


Figure C.3 is suggestive of no overall clear patterns of movement - either towards or away from particular preference values. Table C.1 represents the average changes in each of the preference parameters. The p-values are reported for a null of no change.

Table C.1: Changes in Preferences , non-deliberation/control group

	Own payoff (θ_1)	Joint payoff (θ_2)	Inequality in payoffs (θ_3)
Average change	-0.0385	0.0091	0.0294
Std Error	0.0294	0.0136	0.0278
p-value	0.199	0.508	0.298

References

1. Akerlof, G. A. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics*, 94(4): 749-775.
2. Alpizar, F., F. Carlsson, and O. Johansson-Stenman. 2008. "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica." *Journal of Public Economics*, 92: 1047-1060.
3. Anderson, L. R. and C. A. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review*, 87(5), 847-862.
4. Anderson, J., S. V. Burks, J. Carpenter, L. Gotte, K. Maurer, D. Nosenzo, R. Potter, K. Rocha, and A. Rustichini. 2013. "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples." *Experimental Economics*, 16: 170-189.
5. Andreoni, J. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447-1458.
6. Andreoni, J. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving?" *Economic Journal*, 100(401): 464-477.
7. Andreoni, J. 2006. "Philanthropy." In *Handbook of the Economics of Giving, Altruism and Reciprocity*, S-C. Kolm and J. Mercier Ythier (eds), North Holland.
8. Andreoni, J. and J. H. Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2): 737-753.
9. Asch, S. E. 1952. *Social Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
10. Asch, S. E. 1956. "Studies of Independence and Conformity. A Minority of One Against a Unanimous Majority." *Psychological Monographs*, 70(416):1-70.
11. Bandiera, O., I. Barankay, and I. Rasul. 2005. "Social preferences and the response to incentives: evidence from personnel data." *Quarterly Journal of Economics*, 120(3): 917-962.
12. Banerjee, A. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107(3): 797-817.
13. Bardsley, N. and Sausgruber, R. 2005. "Conformity and reciprocity in public good provision." *Journal of Economic Psychology*, 26(5): 664-681.
14. Baron, R. S., S. I. Hoppe, C. F. Kao, B. Brunzman, B. Linneweh, and D. Rogers. 1996. "Social Corroboration and Opinion Extremity." *Journal of Experimental Social Psychology*, 32: 537-560.
15. Becker, G. 1974. "A theory of social interactions." *Journal of Political Economy*, 82: 1063-1094.
16. Belot, M., R. Duch, and L. Miller. 2015. "A comprehensive comparison of students and non-students in classic experimental games." *CESS Working Paper* DP2015_002.
17. Benz, M. and S. Meier. 2008. "Do people behave in experiments as in the field? - evidence from donations." *Experimental Economics*, 11(3): 268-281.
18. Benabou, R. and J. Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652-1678.

19. Bergstrom, T. C. 1989. "A Fresh Look at the Rotten Kid Theorem and Other Household Mysteries." *Journal of Political Economy*, 97(5): 1138-1159.
20. Bernheim, B. D. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841-877.
21. Birkhchandani, S. D. Hirshleifer and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100(5): 992-1026.
22. Binmore, K. and A. Shaked. 2010. "Experimental economics: Where next?", *Journal of Economic Behavior & Organization*, 73, 87-100
23. Blanco, M., D. Engelmann, and H. T. Normann. 2011. "A Within-subject analysis of other-regarding preferences." *Games and Economic Behavior*, 72: 321-338.
24. Bolton, G. and A. Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166-193.
25. Brock, W. A. and S. N. Durlauf. 2000. "Discrete Choice with Social Interactions." *Review of Economic Studies*, 68: 235-260.
26. Camerer, C. 2003. "Dictator, Ultimatum, and Trust Games." In *Behavioral Game Theory: Experiments in Strategic Interaction*, Camerer, C. (Ed.), Princeton, NJ: Princeton University Press, 43-117.
27. Camerer, C. G. Loewenstein, and M. Rabin. 2004. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.
28. Carlsson, F. O. Johansson-Stenman, and P. K. Nam. 2014. "Social Preferences are stable over long periods of time." *Journal of Public Economics*, 117:104-114.
29. Carpenter, J., S. Burks, and E. Verhoogen. 2005. "Comparing students to workers: the effects of social framing on behavior in distribution games." *Research in Experimental Economics*, 10: 261-290.
30. Carpenter, J. and C. K. Myers. 2010. "Why volunteer? Evidence on the role of altruism, image and incentives." *Journal of Public Economics*, 94(11): 911-920.
31. Carpenter, J. and E. Seki. 2011. "Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay." *Economic Inquiry*, 49(2): 612-630.
32. Cason, T. and V-L. Mui. 1997. "A Laboratory Study in Group Polarisation in the Team Dictator Game." *Economic Journal*, 107(444): 1456-1483.
33. Cason, T. and V-L. Mui. 1998. "Social Influence in Sequential Dictator Game." *Journal of Mathematical Psychology*, 31(5): 467-475.
34. Cialdini, Robert B. 1993. *Influence: Science and Practice*. New York: Harper Collins College Publishers.
35. Cialdini, Robert B. and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology*, 55: 591-621.
36. Charness, G. and M. Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-869.
37. Charness, G. and M. Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspective*, 26(3): 157-176.

38. Chen, Y., F. M. Harper, J. Konstan, and S. Li. 2010. "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens." *American Economic Review*, 100: 1358-1398.
39. Cooper, D. and J. Kagel. 2010. "Other-Regarding Preferences: A Selective Survey of Experimental Results." In *Handbook of Experimental Economics Volume 2*; Kagel, J and A. Roth (eds), Princeton, NJ: Princeton University Press.
40. Deutsch, M. and H. Gerard. 1955. "A Study of Normative and Informational Influences upon Individual Judgement." *Journal of Abnormal and Social Psychology*, 51: 629-636.
41. De Oliverira, A. C. M., R. T. A. Croson, C. Eckel. 2012. "The stability of social preferences in a low-income neighborhood." *Southern Economic Journal*, 79:15-45.
42. Duncan, G. J., J. Boisjoly, M. Kremer, D. M. Levy, and J. Eccles. 2005. "Peer Effects in Drug Use and Sex Among College Students." *Journal of Abnormal Child Psychology*, 33(3): 375-385.
43. Eckel, C. and R. Wilson. 2007. "Social learning in coordination games: does status matter?" *Experimental Economics*, 10(3): 317-329.
44. Engel, C. 2011. "Dictator games: a meta study." *Experimental Economics*, 14: 583-610.
45. Engelman, D. and M. Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review*, 94(4): 857-869.
46. Eyster, E. and M. Rabin. 2014. "Extensive Imitation is Irrational and Harmful." *Quarterly Journal of Economics*, 1861-98, doi: 10.1093/qje/qju021.
47. Falk, A. 2007. "Gift Exchange in the Field." *Econometrica*, 75(5): 1501-1511.
48. Falk, A. S. Meier, and C. Zehnder. 2013. "Do lab experiments misrepresent social preferences? The case of self-selected student samples." *Journal of the European Economic Association*, 11(4): 839-852.
49. Fehr, E. and K. M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.
50. Fehr, E. and Gächter, S. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159-181.
51. Fehr, E. and U. Fischbacher. 2003. "The nature of human altruism." *Nature*, 425:785-791.
52. Fehr, E. and A. Leibbrandt. 2011. "A field study on cooperativeness and impatience in the tragedy of the commons." *Journal of Public Economics*, 95(9): 1144-1155.
53. Fehr, E. and K. Hoff. 2001. "Tastes, Castes, and Culture: The Influence of Society on Preferences", *The Economic Journal*, 121: 396-412.
54. Festinger, Leon. 1954. "A Theory of Social Comparison." *Human Relations*, 7(2): 117-140.
55. Fischbacher, U., S. Gächter, E. Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economic Letters*, 71(3): 397-404.
56. Fischbacher, U. and S. Gächter. 2010. "Social preferences, beliefs and the dynamics of free riding in public good experiments." *American Economic Review*, 100: 541-556.
57. Fisman, R. S. Kariv, and d. Markovits. 2007. "Individual Preferences for Giving." *American Economic Review*, 97(5): 1858-1876.

58. Flinn, C. and D. Del Boca. 1995. "Rationalizing Child-Support Decisions." *American Economic Review*, 85(5):1241-1262.
59. Frey, B. S. and S. Meier. 2004. "Social comparisons and pro-social behavior. Testing 'conditional cooperation' in a field experiment." *American Economic Review*, 94(5): 1717-1722.
60. Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6(3): 347-369.
61. Gächter, S. and E. Fehr. 2002. "Fairness in the Labour Market: A Survey of Experimental Results." In *Survey in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*, F. Bolle and M. Lehmann-Waffenschmidt (eds), Heidelberg: Physica-Verlag.
62. Gächter, S., B. Herrmann, and C. Thoni. 2004. "Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence." *Journal of Economic Behavior & Organization*, 55: 505-531.
63. Gaviria, A. and S. Raphael. 2001. "School-Based Peer Effects and Juvenile Behavior." *Review of Economics and Statistics*, 83(20): 257-268.
64. Glaeser, E., B. Sacerdote, and J. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics*, 111: 507-548.
65. Greiner, B. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114-125.
66. Harbaugh, W. T. 1998. "What do donations buy? A model of philanthropy based on prestige and warm glow." *Journal of Public Economics*, 67: 269-284.
67. Harris, D., B. Herrmann, A. Kontoleon, and J. Newton. 2015. "Is it a norm to favour your own group?" *Experimental Economics*, 18(3): 419-521.
68. Henrich, J., Ensminger, J. McElreath, R., Barr, A., Barret, C., Bolynatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Hencrich, H., Lesogorol, C., Marlowe, F., Tracer, D., and Ziker, J. 2010. "Market, religion, community size and the evolution of fairness and punishment", *Science*, 327: 1480-1484.
69. Henrich, J., Heine, St. J. and Norenzayan, A. 2010. "The weirdest people in the world", *Behavioral and Brain Sciences*, 33: 61-135.
70. Hoffman, E., K. McCabe, and V. Smith. 1996. "Social distance and other-regarding behavior in dictator games." *American Economic Review*, 86(3): 653-660.
71. Huck, S. and J. Oechssler. 2000. "Informational cascades in the laboratory: Do they occur for the right reasons?" *Journal of Economic Psychology*, 21(6): 661-671.
72. Hung, A. and C. Plott. 2001. "Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions." *American Economic Review*, 91(5): 1508-1520.
73. Jones, S. R. G. 1984. *The Economics of Conformism*. Oxford, UK: Basil Blackwell.
74. Karlan, D. S. 2005. "Using experimental economics to measure social capital and predict financial decisions." *American Economic Review*, 95(5): 1688-1699.
75. Karlan, D. S. and List, J. A. 2007. "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment." *American Economic Review*, 97(5): 1774-1793.

76. Kluger, T., E. E. Kausel, and M.G. Kocher. 2012. "Are groups more rational than individuals? A review of interactive decision making in groups." *WIREs Cognitive Science*, 3(4):471-482.
77. Knez, M. and C. Camerer. 1995. "Outside Options and Social Comparison in Three-Player Ultimatum Game Experiment." *Games and Economic Behavior*, 10:65-94.
78. Kohler, H-P. 2001. *Fertility and Social Interaction*. Oxford: Oxford University Press.
79. Krupka, E. L. and Weber, R. A. 2009. "The Focusing and Informational Effects of Norms on Pro-Social Behavior." *Journal of Economic Psychology*, 30(3): 307-320.
80. Laury, S. K. and L. O. Taylor. 2008. "Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good?" *Journal of Economic Behavior & Organization*, 65(1): 9-29.
81. Levitt, S. and J. A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.
82. List, J. A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*, 115(3):482-493.
83. Locke, J. 1693. *Some Thoughts Concerning Education*. London: Printed for A. and J. Churchill.
84. Luhan, W. J., M. G. Kocher, and M. Sutter. 2009. "Group polarization in the team dictator game reconsidered." *Experimental Economics*, 12: 26-41.
85. Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531-542.
86. Manski, C. F. 1995. *Identification Problems in the Social Sciences*. Cambridge, Mass: Harvard University Press.
87. Manski, C. F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives*, 14(3):115-136.
88. Mas, A. and E. Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1): 112-145.
89. McElroy, M. 1990. "The Empirical Content of Nash-Bargained Household Behavior." *Journal of Human Resources*. 25: 559-583.
90. Moscovici, S. and M. Zavalloni. 1969. "The Group as a polarizer of attitudes." *Journal of Personality and Social Psychology*, 12(2): 125-135.
91. Rilling, J. K. and A. G. Sanfey. 2011. "The Neuroscience of Social Decision-Making." *Annual Reviews of Psychology*, 62: 23-48.
92. Rosenzweig, M. and Wolpin, K. 1994. "Parental and Public Transfers to Young Women and Their Children." *American Economic Review*, 84: 1195-1212.
93. Rustagi, D., S. Engel, and M. Kosfeld. 2010. "Conditional cooperation and costly monitoring explain success in forest commons management." *Science*, 330.6006: 961-965.
94. Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116(2): 681-704.

95. Selten, R. 1991. "Properties of a Measure of Predictive Success." *Mathematical Social Sciences*, 21: 153-167.
96. Selten, R. and S. Krischker. 1983. "Comparison of two theories for characteristics function experiments." In R. Tietz (ed), *Aspiration Levels in Bargaining and Economic Decision Making*, Berline and Heidelberg: Springer, 259-264.
97. Shang, J. and R. Croson. 2009. "Field experiments in charitable contribution: the impact of social influence on the voluntary provision of public goods." *Economic Journal*, 199: 1422-1439.
98. Soetevent, A. R. 2005. "Adonymity in Giving in a Natural Context: A Field Experiment in Thirty Churches." *Journal of Public Economics*, 89(11-12): 2301-2323.
99. Stoop, J., C. N. Noussair, and D. Va Soest. 2012. "From the lab to the field: Cooperation among fishermen." *Journal of Political Economy*, 120(6): 1027-1056.
100. Stoop, J. 2014. "From the lab to the field: envelopes, dictators and manners." *Experimental Economics*, 17:304-313.
101. Sunstein, C. R. and R. Hastie. 2015. *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. Boston, MA: Harvard Business Review Press.
102. Tajfel, H., M. Billig, R. Bundy, and C. Flament. 1971. "Social categorization in intergroup behavior." *European Journal of Social Psychology*, 1: 149-178.
103. Tajfel, H. and J. Turner. 1986. "The social identity theory of intergroup behavior." In S. Worchel and L. Austin (eds.). *Psychology of Intergroup Relations*. Chicago: Nelson-Hall.
104. Vesterlund, L. 2003. "The informational value of sequential fundraising." *Journal of Public Economics*, 87(3): 627-657.
105. Vesterlund, L. 2006. "Why do people give?" In W. W. Powell and R. Steinberg (eds), *The nonprofit sector: A research handbook*, vol. 2: 168-190.
106. Young, H. P. 2014. "The Evolution of Social Norms." *Annual Review of Economics*, 7: 359-387.
107. Zimmerman, D. J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*, 85(1): 9-23.