

Good Samaritans and the Market: Experimental Evidence on Other-Regarding Preferences

Michele Belot* Marcel Fafchamps†

April 2013

Abstract

This paper builds on the evidence showing that people behave pro-socially in small groups or partnerships but are more selfish in market-like situations. We construct an experiment to study what preferences people exhibit when searching for a partner, that is, whether they apply market-like heuristics or whether they behave more pro-socially as they typically do once these partnerships are formed. In our market-like situation, people need to select a type of partner, either high or low. The payoffs of both players depend on the types of partners. Decisions are implemented unilaterally and effectively affect three other people in the economy, since the selected partner can no longer team up with others who thus must partner with each other. We compare this situation to two alternative situations where people are first assigned to smaller groups of four consisting of two high and two low participants. In one case, they are asked again to choose between a high and low partner. In the other case, they are simply asked to choose between two payoff allocations for group members. These two situations increase the saliency of the implications of decisions for others and for efficiency. We find that a market-like situation reduces the ‘good samaritan’ spirit in this environment as well: when choosing a partner agents are less likely to sacrifice their own material well-being to increase the well-being of others.

*Department of Economics, University of Edinburgh and CESS, Nuffield, Oxford University

†Department of Economics, Oxford University

1 Motivation

Partnership formation is at the core of economic and social life: friendships, marriages, business partnerships or political alliances are a fundamental underpinning of economic and social exchange. In turn, this network of partnerships and social ties sustains societal divides along many socio-economic dimensions such as race, religion or education (McPherson, 2001).

Very often partnerships are formed voluntarily, and arise from a process that entails choosing one another and agreeing on how the gains from partnership are shared. There is many examples of pairings of two types of agents – e.g., producers and investors, farmers and landowners, advisors and advisees – in which collaboration is voluntary and entail a sharing agreement. The sorting patterns resulting from such partnership formation play a crucial role in the allocation of resources within a society, both in terms of efficiency and equality. Economists and sociologists have long been interested in how relationships are formed and who matches with whom, starting with the early work by Becker (1973).

This paper studies the role of other-regarding preferences in the process of partnership formation. The literature on decentralized matching is agnostic as to what role these preferences may play. But there is now a well established literature on social preferences (Fehr and Schmidt 1999, Bolton and Ockenfels 2006, Charness and Rabin 2002) showing that people are willing to transfer resources to others, in particular if these transfers reduce inequality, and that the role of other-regarding preferences seems particularly salient in the context of small partnerships. Take the example of household production. Without a modicum of altruism, free-riding would limit the provision of household public goods, making households less efficient and, hence, less likely to survive as production and consumption units. Given how important household stability must have been for reproductive success over the course of human history, we expect human societies to have developed strong norms and behavioral traits that favor other-regarding behavior within small groups (e.g., Fehr and Schmidt 1999, Ledyard, 1995, Camerer, 2003, Henrich et al. 2005).

In contrast, exchange in competitive markets can deliver an efficient outcome even if all partici-

pants act to maximize their own material welfare and nothing else. The original insight goes back to Adam Smith's shoemaker parable, and it has been verified in numerous market experiments (e.g., Smith 1962). Indeed, in a review of the experimental evidence, Bowles (1998) observes that the more the experimental situation approximates a competitive (and complete contracts) market with many anonymous buyers and sellers, the less other-regarding behavior is observed. This finding fits with the two apparently opposite views of Adam Smith who argues in the *Wealth of Nations* (1776) that self-interest prevails in markets, while acknowledging the pro-sociality of human nature in the *Theory of Moral Sentiments* (1759). Vernon Smith (1998) points out that these two forms of human nature are not contradictory but apply to different contexts. Selfish behavior maximizes the gains from impersonal market exchange, while cooperative behavior maximizes the gains from non-market personal exchange.

So far, little is known as to what role do other-regarding preferences play in partnership formation. The literature on decentralized matching describes the process of match formation as a market-like process while the literature on other-regarding preferences suggests that such preferences are particularly strong in small partnerships. So the question is: do people apply market-like heuristics when searching for a partner (i.e. behave selfishly); or do they behave more pro-socially, as they do once these partnerships or small entities are formed? And if they do behave differently, what motivates differences in behavior?

This paper provides experimental evidence of the role of other-regarding preferences in the process of partnership formation and sorting. We focus on one possible mechanism that distinguishes markets from personal relations, namely, how salient are the implication of one's choice on others' payoff. In a small group situation, the implications of decisions on others are clear and salient. But they are much less so in a market situation. Hence participants may feel empowered to pursue their own material welfare and ignore the consequences of their choice on others. To illustrate, imagine one remaining bread and several consumers. In a small group allocation, it is immediately obvious that taking the last bread has implications on others. People may therefore feel morally or socially obliged to share or let others have the remaining bread. In contrast, if the

bread is the last one at the shop, the implications of taking it on others are the same – both they are less salient. We suspect that many people will take the last bread if they can, and will feel little guilt for depriving others from it.

We consider different types of other-regarding preferences: pro-social preferences (Fehr and Schmidt 1995; Bolton and Ockenfels 2006); invidious preferences (Blanchflower and Oswald 2004; Fafchamps and Shilpi 2008); and preferences for efficiency (Charness and Rabin 2002). In our design we separately vary the implication and saliency of partner choice on efficiency and income distribution. We study unilateral partnership formation - i.e., partnership formation without mutual consent – in order to identify preferences separately from strategic considerations in partnership formation.

We consider an environment with two "categories" of people intended to represent, in an un-contextualized way, the two sides of a partner selection choice. 24 participants are invited to participate to each session. They are assigned to one category, and to one of two types – ‘high’ and ‘low’ – based on their performance in a real effort task before the experiment proper begins. The first treatment we introduce is a simple small group dictator allocation decision involving groups of four people (two high and two low types from each category). Participants are asked to choose between two divisions of payoffs among the four, with no reference to partner selection. In a second treatment, we present the same allocation problem, but the decision is framed as a choice between a high or low partner. In the third treatment the allocation problem is also framed as a choice of partner, but unlike in treatment 2 the decision is put in the context of a large "market" with many participants. In the two latter treatments, experimental subjects are shown the payoff distributions associated with all four types of partnerships. In all three cases, the decision of one of four players is selected at random and implemented in a unilateral manner to determine the payoff of all four. This allows us to abstract from strategic considerations when we infer preferences from choices. But the treatments vary the salience of the effects of one’s choices on others. In the first treatment, the implications that decisions have on others are obvious and salient. They become less salient as we move towards a more market-like situation. We conjecture that this decrease in saliency may

trigger selfish heuristics.

In our three experimental treatments we find that participants mostly follow their material self-interest. This is in line with numerous studies of behavior in market games. We find a significant difference in partner selection between large groups (treatment 3) and small groups (treatment 2), but a much smaller difference between treatment 2 and the group treatment 1. This suggests that participants display other regarding preferences whenever they perceive their decision has a direct impact on others. In treatment 3 they are less likely to sacrifice their individual payoff to increase aggregate efficiency. As a result aggregate efficiency falls since the design rules out the operation of competitive forces. There is little or no statistical difference between treatments 1 and 2.

Because we experimentally assign participants to different average payoffs, we can test whether deviation from pure selfish behavior varies with expected income from the experiment. We find that high payoff agents are on average more altruistic than low payoff agents, but significantly less so in a partner selection environment. Similarly, low payoff agents are less reluctant to reduce others' payoff in a partner selection environment, particularly if the number of participants is large. We interpret these findings as suggesting that a large partner-selection environment reduces the 'good samaritan' spirit: agents are less likely to sacrifice their own material well-being to increase the well-being of others.

The paper is structured as follows. Section 2 presents the experimental design and the different treatments. The testing strategy is outlined in Section 3. Empirical results are presented in Sections 4 and 5.

2 Experimental Design

Participants play in sessions of 24¹ and participants in a session only play one treatment – i.e., we use a 'between subjects' design. A detailed description of the experimental protocol is given in Appendix.

¹In 3 sessions (one for each treatment), the number of people who reported for the experiment was less than 24. Hence the experiment was played in groups of 20 instead. This does not affect the experiment except in treatment 3.

2.1 Stages

Within each session the experiment is divided into two stages. In the first stage, the pool of participants is divided equally and randomly into two categories A and M .² The two categories are intended to represent, in an uncontextualized way, the two sides of a matching game, e.g., bride-groom, employer-employee, or hospital-intern. Having been assigned to a category, each participant individually completes a computerized task. Subjects are asked to do simple calculations for a period of 3 minutes – additions or multiplications depending on the category, A or M , to which they have been randomly allocated. Based on their performance in the task relative to other participants in the same session and category, they are assigned one of two types – bottom 50% or top 50%. Here we denote the two types simply as ‘high’ and ‘low’.

In the second stage of the experiment, participants play six rounds of an allocation game. No feedback is provided to participants until the end of the experiment, at which time they are only told their final payoff. Since payoffs are based on one randomly selected round, it is impossible for participants to infer the choices of other participants. The purpose of this is to rule out repeated games and strategic play: each round is de facto a dictator game with anonymous others and no feedback.

In the first two treatments, participants are assigned to groups of four. Each group consists of a low and high type participant from each category. In treatment 1 (T1) choices are presented as a selection between two ‘pies’ divided into four possibly unequal slices – see Instructions in Appendix A. For each round, each participant in each group selects his favorite pie/allocation. At the end of the session, one subject from each group and one round are selected at random, and his/her choice determines the payoff of all four participants in the group. The structure of the game thus resembles a dictator game with four players.

In treatment 2 (T2) participants are asked to indicate whether they would prefer to form a partnership with the low or high type participant from the other category. They are shown how the payoffs would be distributed for each possible partnership (high A /high M , high A /low M , low

²In the experiment these categories are refereed to as ‘Addition’ and ‘Multiplication’ – see below.

$A/\text{low } M, \text{ low } A/\text{high } M$). They are also made aware of the implications of their choice for the 2 other participants in their group. For example, if a participant in category M , say, selects the high type in category A as partner, this also determines the payoff of the two remaining players since they can now only be matched with each other. Participants are made aware of this – see Figure A2. As in treatment 1, final payoffs are determined by randomly selecting one round and one subject from each group, and letting his/her choice of partner determine the payoff of all four participants in the group.

In treatment 3 (T3), all 24 participants in a session play a partner selection game together as follows. Each participant is asked to select one of two possible types of partners, high or low, among players of the other category. Since the 24 participants are divided equally between categories A and M , and subsequently divided equally between low and high type within each category, there are six participants in each category \times type. Because players are anonymous and there is no feedback, the choice of each player resembles treatment 2 except for the larger number of players. They are also shown how the payoffs would be distributed for each possible partnership, as in treatment 2. But they are not told anything explicitly about the implications of their choice for others. To understand these implications, they need to understand that by choosing a partner of a certain type, they prevent one person from their own category to be matched to a partner of that particular type.

Payoffs in treatment 3 are determined as follows. We first randomly select one category (A or M) at random as well as one of six rounds. We then aggregate the choices of the participants in the selected category. If there is excess demand for one type, then the scarce type is allocated in a random manner between those who have expressed a preference for it. For instance, say category A is selected. Of the 12 participants in category A , 8 have selected to match with a high type. Since there are only 6 high types in category M , each of the 8 participants is allocated to a high type match with a probability equal to $6/8$, and a low type match with probability $2/8$. The 4 participants who have selected to match with a low type get their choice of type. The payoff determination process is explained in detail to participants before the experiment (see Instructions

in Appendix for details).

The three treatments differ in the salience with which a subject's choice affects other players. In treatment 1, it is obvious by design that selecting one pie affects one's payoff and that of three other players. In treatment 2, it is clear to each player that their choice affects the payoff of the partner they choose. Given the payoff determination rule, they can also deduce that selecting one partner de facto forces the other two players together. Alternatively they may follow a market logic and convince themselves that their choice directly affects only one other player, and hence that their other-regarding preferences only apply to that player.

In treatment 3 players can, as in treatment 2, clearly see the effect of their choice on their – yet to be determined – *partner*. But by taking one possible partner away from the choice set of other participants, they also de facto limit the choices of other players and impose a specific match to two other – yet to be determined – participants. This latter effect is less salient than in treatment 2. Given that all participants are anonymous to each other and never receive feedback on each other's choices, treatment 2 and 3 are ultimately equivalent: ego's choice affects the payoff of three other players. But realizing this takes some sophistication, which we formalize later in this paper. Because of the various rounds of randomization that take place before payoffs are assigned, treatment 3 may blur the sense of responsibility that participants associate with their actions. We call this the 'dilution hypothesis'. A formal presentation of this hypothesis is given Appendix 2.

2.2 Payoffs

Payoffs in the second stage of treatments 2 and 3 represent how gains from matching are shared between matched partners. Gains are always shared equally if both partners are of the same type (both high or both low). If partners are of different types, the division of payoffs differs from game to game so as to vary the efficiency and equity of an heterogeneous match. We call each payoff matrix a scenario.

In a session all participants play 6 different scenarios. The scenarios are the same for all the participants in the same session. The same set of scenarios/payoff matrices is used for the three

treatments. We use a set of 17 different scenarios. In some scenarios, payoffs are such that an heterogeneous match – and hence negative assorting (e.g., low A with high M) – is more efficient; in other scenarios, payoffs are such that positive assorting is efficient. We also vary how the gains from a match are divided between the two participants. The three different division rules used are summarized in Table 1.

The scenarios are summarized in Table 2 where are reported the payoffs associated with the different types of partnerships (or different allocations in treatment 1). The different scenarios have been chosen so that we can assign as unambiguously as possible a sequence of choices made by an individual subject to a specific preference archetype, provided that the subject’s choices are all consistent with that archetype. We come back to this in Section 3.4.

Table 1. Sharing rule for heterogeneous matches

	Low A	High A
		1/2, 1/2
Low M	1/2, 1/2	1/3, 2/3
		1/6, 5/6
	1/2, 1/2	
High M	2/3, 1/3	1/2, 1/2
	5/6, 1/6	

High types have higher payoffs on average, and low types low payoffs. To help identify other-regarding preferences,³ we introduce a further payoff differentiation between low A and low M payoffs such that low M types get, on average, lower payoffs than low A types. Throughout the analysis we present results broken down by these three payoff categories.

It is important to realize that in each T2 and T3 scenario participants ultimately choose between two possible pairings: (1) high A -high M / low A -low M and (2) high A -low M / low A - high M . The first corresponds to positive assorting, the second to negative assorting. Each of the pairings has an associated total payoff for the four participants affected by someone’s choice. Hence to each scenario is associated two possible efficiency values. In some scenarios positive assorting is efficient; in others negative assorting is efficient. Treatment T1 mimics these differences albeit without explicit assorting. In all but 3.8% of the observations one form of assorting is more efficient than

³Specifically, we need this design to distinguish between selfish preferences and maximin preferences for low types.

the other. However, whether positive or negative assorting maximizes someone’s material payoff varies by category and type and according to the sharing rule associated with a particular scenario (see Table 1).

Treatments may affect expressed preferences in different ways. Because treatment 1 resembles the allocation process that takes place within a household, where some members (e.g., the person in charge of grocery shopping) make choices for other members, it may elicit preferences that are more other-regarding. One possibility is that participants pay more attention to the efficiency cost of their action. Another is that they seek a more equal allocation of payoff, or that they are less spiteful.

Because treatment 3 resembles an anonymous market environment, participants may use market heuristics when making choices, feeling empowered to pursue their own material welfare and thereby ignoring the consequences of their choice on others. Treatments 2 and 3 also raise the possibility that players may prefer to match with someone of their type (e.g., high or low) – see for instance Curriani and Mengel (2011) for recent evidence of homophily in experiments with randomly assigned types.

3 Analysis and results

The experiment took place at the experimental laboratory of the Centre for Experimental Social Sciences in June and November 2011.⁴ We ran 12 sessions (4 sessions per treatment). 308 participants took part in total. Participants earned £10.28 on average.

The first objective of the experiment is to test whether play is systematically different between treatments and, in particular, whether decisions are more sensitive to (1) the payoffs of others and (2) to efficiency considerations as the saliency of the implications of choices increases. The second objective is to test what type of preference is most likely to account for observed behavior. We report on these two objectives in turn.

⁴We ran the sessions corresponding to Treatment 3 in June 2011 and those corresponding to Treatments 1 and 2 in November 2011.

3.1 Sensitivity to the payoffs of others

Throughout we denote by π_i the payoff of subject i and N_i be the set of four players affected by i ' choice.⁵ We define efficiency as the sum of the payoffs of the four participants affected by the choice of a single player.

3.1.1 Variation across treatments

We begin by showing that play varies systematically with treatments. In Table 3 we report, for each treatment, the proportion of individual choices that maximize one's own material payoff. From the first panel of the Table, we see that most choices are those that maximize the subject's material payoff. There is little difference across treatments, and whatever differences are present are not statistically significant. We also note that the proportion of choices consistent with selfish preferences is highest for low average payoff participants and lowest for high average payoff participants.

The second panel of the Table shows the proportion of individual choices that maximize the aggregate payoff of the other three players affected by ego's choice. Choices in which own payoff or other's payoff is unaffected by i 's choice are omitted from the calculation. The proportions are reported separately for high type players, low type players in category A and low type players in category M . The reason is that these players face different individual payoffs and thus different choices. We also report a χ^2 test of equality of means with its p -value below.

The other three panels of Table 3 show the proportion of choices that maximize the payoff of the other three players affected by i 's selection. The top right panel presents the average over all choices. We note that the proportion of choices that maximizes others' payoffs decreases systematically as we go from treatment T1 to treatment T3. This is true for all payoff types, but is strongest and highly statistically significant for high payoff participants. For the lowest payoff types, the difference is not statistically significant at standard levels. For the intermediate payoff category, the difference is significant at the 10% level. This suggests that participants take others' payoff less

⁵In treatment 3, the exact individuals affected are unknown prior to ex post randomization, but their payoffs are known since, by design, they belong to the four category \times type groups. See appendix for a formal demonstration.

into consideration as we move from treatment T1 to treatments T2 and T3. We also note that the magnitude of the increase is similar between treatments T1 and T2 and between T2 and T3.

In the lower half of Table 3 we split choices that maximize others' payoff depending on whether doing so also maximizes one's own payoff or not. The 'good Samaritan' spirit correspond to the case where participants maximize others' payoff even though doing so reduces their own (lower right panel of Table 3). The other panel is consistent with efficiency consideration but is less remarkable from an equity point of view. From the lower left-hand panel, we find that participants are more sensitive to aggregate efficiency in treatment T1 than in partner-selection treatments T2 and T3. This is true for all three payoff categories of participants, but only significantly so for high average payoff participants. From the lower right-hand panel, we see that the good Samaritan spirit is highest in treatment T1 and lowest in T3. The difference is large in magnitude and significant at the 1% level for high and middle average payoff participants; for low payoff participants the difference is smaller in magnitude and only significant at the 16% level. That is, low payoff participants tend to maximize their own payoff and show less regard for the payoff of others, suggesting that they are perhaps trying to make up for having been assigned to the lowest payoff category at the onset of the experiment.

To further verify our results, we investigate whether treatments affect the sensitivity of participants' choice to differences in payoffs between the two options they face. Let $\Delta\pi_i \equiv \pi_i^h - \pi_i^l$ be the gain in i 's payoff from choosing a high partner (or equivalent allocation in T1). Similarly let $\Delta\pi_{-i} \equiv \sum_{j \neq i, j \in N_i} (\pi_j^{h_i} - \pi_j^{l_i})$ be the gain in the payoff of the other three players affected by i 's choice. To calculate the marginal effects of $\Delta\pi_i$ and $\Delta\pi_{-i}$ on the probability of choosing a high partner (or equivalent allocation in T1) we estimate a regression model of the form:

$$y_i = \Delta\pi_i \otimes T \otimes X + \Delta\pi_{-i} \otimes T \otimes X + \varepsilon_i \quad (1)$$

where $\Delta\pi_i \otimes T \otimes X$ is shorthand for all the possible interaction terms between them and similarly for $\Delta\pi_{-i} \otimes T \otimes X$.⁶ Regression (1) is estimated using a linear probability model with standard

⁶In other words, we include terms in $\pi_i, T, X, TX, \pi_i T, \pi_i X$ and $\pi_i TX$ where T and X are themselves vectors of dummies. Similarly for π_{-i}, T and X .

errors clustered at the session level.

We report in Table 4 the estimated marginal effects with their t -value for each of the $T \otimes X$ combinations. What the Table reveals is that choice sensitivity to the payoff difference between the two options is comparable across treatments: a one unit increase in payoff gain increases the probability of choosing the more beneficial option by 5 to 6 percentage points across all subject types and treatments. Since the standard deviation of $\Delta\pi_i$ is 4.67, this is a large effect. There is a decrease in sensitivity to own payoff in treatment T1 for high types and low A types, but the difference is relatively small in magnitude, not statistically significant, and we observe nothing similar for low M types.

In contrast, choice sensitivity to $\Delta\pi_{-i}$ varies dramatically across treatments and subject types. For high types – who on average earn higher payoffs – sensitivity to $\Delta\pi_{-i}$ is absent in treatment T3 but present in the other two treatments. The effect is large in magnitude: a one standard deviation (i.e., 7.67) increase in $\Delta\pi_{-i}$ raises the probability of choosing high by 7.5% in treatment T2 and 11.8% in treatment T1. In treatment T3 the effect is numerically 0. This suggests that among these players considerations of altruism or efficiency are eliminated in T3, a finding that is consistent with the dilution hypothesis.

Results are different for low types. Here we find that, under anonymous partner selection T3, low type participants are at the margin *less* likely to choose a high partner if other players benefit more from that choice, controlling for their own payoff gain. The effect is large in magnitude, especially for low A types: a one standard deviation increase in $\Delta\pi_{-i}$ reduces the probability of choosing high by 11.4% for low A and 7.1% for low M types.

These findings are to be read in the context of the literature on inequality aversion (e.g., Fehr and Schmidt 1999, Okada and Reidl 2005). Experimental evidence has suggested that individuals display a desire to reduce the difference between their payoff and that of others both from above and from below. In other words, if a subject has a high payoff relative to other participants, this subject is often observed taking redistributive actions that reduce the difference between her payoff and that of other participants. This is consistent with the behavior of high types in our experiment,

who have higher average payoffs and, in treatments T2 and T1, are more likely to choose an action that increases the payoff of other participants who, on average, earn a lower payoff. In other words, high types often choose an action that reduces the difference between their payoff and the lower payoff of others.

Also according to inequality aversion, a subject who has a low payoff relative to others often takes actions that increase her payoff at the expense of others. This is what we observe low types do in treatment 3: controlling for their own payoff, low types – who on average earn lower payoffs – are more likely to take an action that reduces the payoff of others. What is interesting is that, in our experiment, the two behaviors do not coexist: altruism (inequality aversion from above) is only present in treatments that emphasize the effect one’s choice has on several others; envy or spite (inequality aversion from below) is only present in the treatment that blurs the effect of one’s choice on others. This suggests that the dilution effect reduces altruism/efficiency considerations, but not envy which is, rather, exacerbated by anonymity. In other words, the two sides of inequality aversion respond differentially to an anonymous market/partner selection environment: altruism is blunted by it, while envy is increased. In our experiment, the two effects together combine to reduce the efficiency of participants’ choices.

3.1.2 Choices and efficiency

We next investigate the relationship between treatment and the efficiency of individual choices. We ignore the 60 observations in which choices generate the same total payoff for the four players affected by ego’s choice. We focus on whether the choice ego makes is efficient or not. In treatment T1 participants choose the efficient allocation in 70% of the observations. This proportion falls to 67% in T2 and 61% in T3. This difference is statistically significant at the 1% level.

We reproduce this finding in the first column of Table 4. Treatment T3 is the default category so that reported coefficients capture the efficiency gain of individual choices in T2 and T1 relative to T3. Furthermore, if we test average efficiency between T1 and T2, the difference is not statistically significant. In other words, participants are more likely to opt for an efficient allocation when the

effect they have on others' payoff is more salient. In contrast, individual choices are significantly less efficient in an anonymous partner selection setting.

In the other columns of Table 4 we disaggregate the result and regress the efficient choice dummy on treatment for each of the three payoff categories separately. The results indicate that the T1 and T2 are associated with more efficient choices for both high and middle payoff participants. For low payoff participants, the treatment effect is not significant although, in terms of average efficiency across all treatments, low payoff participants are not statistically different from other payoff categories. It appears that an anonymous partner selection setting enables higher payoff participants to behave in a more selfish manner while in treatments T1 and T2 they may feel moral pressure to behave altruistically towards lower payoff participants. In contrast, low payoff participants do not seem to have this concern, perhaps because they feel more entitled to pursue their self-interest, having been assigned to the lowest payoff category at the onset of the experiment.

3.2 Preference archetypes

In this subsection we adopt a more structural approach and assign participants to archetypes summarizing the form of their other-regarding preferences. We test whether this assignment varies with treatment. The idea behind the approach is that people behave in ways that may vary with the decision context, and that their behavior can be approximated by preference archetypes. If we know what archetypes best capture the behavior of a large fraction of the population in a given environment, we may be in a better position to predict what types of behavior to expect in that environment.

Scenarios were designed to facilitate the assignment of participants to the six different archetypes listed below.

1. *Selfish*: A selfish player is defined as someone who maximizes π_i .
2. *Efficient*: An efficient player is someone who maximizes $\bar{\pi}_i = \sum_{j \in N_i} \pi_j$.
3. *Equity only*: Choose the allocation that minimize absolute inequality defined as $\sum_{j \in N_i} |\pi_j - \bar{\pi}_i|$

4. *Spiteful*: Choose the allocation that maximizes one’s relative payoff $\pi_i - \frac{1}{3} \sum_{j \neq i, j \in N_i} \pi_j$
5. *Maximin*: Choose the allocation that maximizes the minimum payoff among the four affected individuals $\max \min \pi_j, j \in N_i$.
6. *Homophily*: Choose a partner from one’s own type (high or low)

We seek to assign each participant to the archetype that best describe their behavior in the six rounds. Different participants may follow different archetypes. We start by identifying participants who follow a single archetype perfectly over all six rounds, and we observe what proportion of participants we can assign in this manner. For high payoff types, payoffs are such that assignment is typically unambiguous. For low types, the choices made by a given subject may be consistent with more than one archetype, depending on the set of six scenarios they faced. Next, we then introduce the possibility of deviations from the behavior predicted by a single archetype. At the end of the section, we allow for the possibility of hybrid preferences.

In Table 5 we present the result of our first calculation. For each archetype k , we calculate $\Delta u_i^k \equiv u_k(\pi_i^h) - u_k(\pi_i^l)$ where preference function $u_k(\cdot)$ is that corresponding to archetype k . For each subject we then count the proportion of rounds (out of a maximum of 6) for which the subject behaves in accordance to archetype k , that is, for which $y_i = 1$ if $\Delta u_i^k > 0$ and $y_i = 0$ if $\Delta u_i^k < 0$. We ignore cases in which $\Delta u_i^k = 0$ because they are uninformative. We say subject i makes choices consistent with archetype k if this proportion is 100%. Depending on scenarios and player type, choices made over six rounds may be consistent with more than one archetype.

Results indicate that the archetype most consistent with observed choices is the ‘selfish’ utility model in which people only care about their material payoff. We do, however, observe systematic differences across treatments and types. Fewer participants are assigned to the selfish archetype in treatment T1 than in other treatments. The difference is strongest for high types, confirming earlier results that suggest these participants behave in a less selfish manner when the experiment is framed as an allocation process rather than as partner selection. Low types tend to behave more in accordance to the selfish archetype than high types, again confirming earlier results.

Turning to other archetypes, we note that very few participants are assigned to the efficiency only archetype, that is, have a utility function that would lead them to always choose the most efficient allocation. We do, however, note that the efficiency only archetype is more common in treatment T1, and among high types. This is consistent with some of our earlier findings that suggest efficiency concerns among some participants, but it reminds us that ‘efficiency only’ is not a suitable description of the way participants behave on average.

A small proportion of participants behave in agreement with the equity only archetype. This archetype assumes that participants always choose the allocation that results in the smallest level of payoff inequality. This archetype is found mostly among low types for whom this overwhelmingly coincides with self-interest (100% overlap in choices between the two archetypes for low A types and 91% overlap for low M types).

A larger proportion of participants fit the invidious or spiteful archetype, that is, make choices that minimize the difference between their payoff and that of others. This archetype is found most often among participants to treatment T3, in line with earlier results emphasizing that the anonymous partner selection treatment seems to encourage invidious choices relative to the other two treatments. Caution is however warranted because 89% of the low M types whose decisions fit the invidious archetype also fit the selfish archetype.

A non-negligible proportion of low A types follows maximin preferences, i.e., they make choices that maximize the minimum payoff to any player. Since low A types nearly always are those players getting the minimum payoff, the data shows that maximin always coincides with self-interest for low A types. Overlap is also common (90%) among low M types. Hence this finding should not be given too much weight. Finally, contrary to other experiments that have documented that participants naturally assort according to randomly assigned type (e.g., Curriani and Mengel 2011), we find very little evidence of homophily in our data: very few participants systematically choose a partner of their type, i.e., low with low or high with high.

At the bottom of the table we report the proportion of participants who were assigned to multiple archetypes, and those whose behavior does not fit any. The last panel of Table 5 confirms

that, by experimental design, high types can nearly always be unambiguously assigned to a single archetype. We were unable to do the same for low types because, with six rounds and four players, there were not enough degrees of freedom in the experiment to design scenarios that are sufficiently informative for both high and low types. This too is confirmed in Table 5. Multiple assignment is most common among low A types, a point that has already been touched upon in the earlier presentation.

The last row of Table 5 indicates that a sizeable proportion of participants do not fit any of the six archetypes we considered. Lack of fit is most noticeable in treatment T1, and this is true for all payoff categories. The behavior of participants in this treatment is less well explained by the simple preference models we considered. But in all treatments there is a sizeable minority of participants whose behavior is not consistent with any of the archetypes we considered.

One possible explanation for this is that people make mistakes: they may have preferences that follow one of our archetypes but, due to inattention or lack of interest, they sometimes make choices that do not correspond to their underlying preferences. To investigate this possibility, we estimate a mixed maximum likelihood model. The starting point of the estimation methodology is the observation that:

$$\Pr(y_i = 1 | \pi_i, \pi_{j \neq i}, T_i) = \sum_{k=1}^K \Pr(y_i = 1 | \pi_i, \pi_{j \neq i}, T_i, k) \Pr(u = u_k | T_i) \quad (2)$$

where π_i and $\pi_{j \neq i}$ denote the four payoffs potentially entering the preference utility u_k of archetype $k = \{1, \dots, K\}$. Since π_i and $\pi_{j \neq i}$ are randomly assigned in the experiment, we can ignore correlation between payoffs and preferences. But we allow preferences to differ across treatments, hence the conditioning on T_i . A similar probability can be derived for $y_i = 0$. Since, for a given treatment T_i , $\Pr(u = u_k | T_i)$ is a constant, we denote it as γ_{kT} .

Next we assume that the probability of choosing action $y_i = 1$ increases in Δu_i^k , the utility gain from choosing $y_i = 1$ that is associated with payoffs π_i and $\pi_{j \neq i}$ when preferences are given by $u_k(\cdot)$. To formalize this idea, we borrow from Luce (1959) and write:

$$\Pr(y_i = 1 | \pi_i, \pi_{j \neq i}, T_i, k) = \frac{e^{\sigma_T \Delta u_i^k}}{1 + e^{\sigma_T \Delta u_i^k}} \quad (3)$$

Parameter σ_T captures how sensitive decisions are to Δu_i^k in treatment T . If $\sigma_T = 0$, the choice between $y_i = 0$ and $y_i = 1$ is random and does not depend on payoffs. If σ_T is arbitrarily large, expression (3) tends to 1 if $\Delta u_i^k > 0$ and to 0 if $\Delta u_i^k < 0$ – which corresponds to the case where choices are perfectly predictable once we know someone’s archetype and the payoffs they face. Intermediate values of σ_T capture situations in which participants systematically diverge from random play in the direction predicted by archetype k . The model assumes that participants are more likely to take the decision predicted by their archetype the larger the utility gain Δu_i^k is between the two choices. In other words, participants make more mistakes when the difference in payoff is small. Mixed models of this kind have successfully been fitted to experimental data (e.g., Andersen et al. 2008, Null 2012).

The likelihood function has the form:

$$L(\gamma_k, \sigma | y_i, \{\Delta u_i^k\}, T_i) = \sum_{k=1}^K \gamma_{kT} \left(\frac{e^{\sigma_T \Delta u_i^k}}{1 + e^{\sigma_T \Delta u_i^k}} y_i + \frac{1}{1 + e^{\sigma_T \Delta u_i^k}} (1 - y_i) \right) \quad (4)$$

and is estimated separately for each treatment, ensuring that $0 < \gamma_{kT} < 1$ and imposing that $\sum_{k=1}^K \gamma_{kT} = 1$.⁷ Once σ_T and $\{\gamma_{1T}, \dots, \gamma_{KT}\}$ have been estimated, we compute, for each subject, the posterior probability that their choices follows a particular archetype.⁸ Accumulating across rounds for each individual i , we get the posterior probability that i follows archetype k :

$$\Pr(k | \{y_i\}) = \frac{\Pr(k) \Pr(\{y_i\} | k)}{\Pr(\{y_i\})} \quad (5)$$

where $\{y_i\} = \{y_i^1, y_i^2, \dots, y_i^6\}$ is the set of decisions made by i over the six rounds. Since $\hat{\gamma}_{kT}$ and $\hat{\sigma}_T$ vary across treatment, $\Pr(k | \{y\})$ also varies across treatment for the same set of choices made.

⁷Estimation is achieved by numerical optimization in Stata. To ensure that all Δu_i^k have the same weight in the estimation, we normalize them to all have a unit standard deviation. Convergence difficulties can arise when $\gamma_{kT} \approx 0$ for some k , or when the Δu_i^k are too correlated across archetypes (akin to multicollinearity).

⁸Let $\Pr(k | y_i = 1)$ denote the probability that individual i is of archetype k if he/she sets $y = 1$. The starting point of our calculation is the following relationship that holds for each choice i makes:

$$\begin{aligned} \Pr(k | y = 1) &= \frac{\Pr(k) \Pr(y = 1 | k)}{\Pr(y = 1)} \\ &= \frac{\Pr(k) \Pr(y = 1 | k)}{\sum_k \Pr(k) \Pr(y = 1 | k)} \end{aligned}$$

For simplicity of exposition, we have omitted the dependence on π and T . Unconditional probability $\Pr(k)$ is estimated by $\hat{\gamma}_{kT}$ while $\Pr(y = 1 | k)$ is obtained from expression (3) using estimated $\hat{\sigma}_T$.

Once we have $\Pr(k|\{y\})$ for each subject, we look at how accurate the predictions are for different individuals, i.e., how accurately they are estimated to follow a given archetype.

Estimates of posterior probabilities are summarized in Table 6. The first panel of the Table reports the average of estimated posterior probabilities (5) calculated using parameters $\hat{\gamma}_{kT}$ and $\hat{\sigma}_T$ estimated using (4). The second panel of the Table assigns each individual to an archetype if their $\Pr(k|\{y_i\})$ exceeds 0.5 for one k – which can happen at most for one k , but could happen for none. As it turns out, all our participants are assigned to one archetype, which suggests that the method was able to infer everyone’s type with accuracy.⁹ We do, however, note that the estimated value of $\hat{\sigma}_T$ is smallest for treatment T3 ($\hat{\sigma}_{T_3} = 6.60$ with a t -value of 5.59), intermediate for T2 ($\hat{\sigma}_{T_2} = 14.48$ with a t -value of 4.25) and largest for T1 ($\hat{\sigma}_{T_1} = 255.29$ with a poorly estimated t -value). This suggests that participants fit one of the six archetypes better in T3 than in T1. This confirms that we are less able to predict individual behavior in T1, perhaps because participants’ choices reflect preferences other than our six archetypes – more about this later.

We find that most participants fit the selfish archetype best. This is true for all treatments and all player types. Secondly, high types are systematically more likely to act selfish in partner selection treatments, particularly T3. This confirms earlier results. Third, participants with the lowest average payoff (low M types) overwhelmingly play selfish, even in the pie allocation treatment. This too is consistent with earlier observations. Fourth, the proportion of players classified as maximizing efficiency increases as we move from treatment T3 to T1. This is particularly noticeable for high types, mirroring earlier observations. But only a small proportion of players are best described as following this archetype.

Fifth, a number of players are classified as having spiteful preferences, that is, as choosing strategies that reduce the difference between their payoff and that of others even if it means reducing their own payoff. (If they had chosen to maximize their own payoff, they would have been classified as selfish, not spiteful.) The proportion of spiteful is highest in treatment T3, especially among low

⁹In fairness, we should point out that we do not correct predicted archetypes for sampling error in parameter estimates. But given how strong assignment is, introducing the correction, if it were possible, would not greatly modify our results.

types. This findings is also in agreement with earlier results.

Finally, we find that a number of low types, especially those in the intermediate payoff category A , are classified as maximin players. This happens more frequently in treatment T1, suggesting that players in that treatment have some equity concerns, but these are not adequately captured by the equity only archetype – which only fits a small number of participants overall, and none of them in treatment T1. Equal pie sharing thus does not appear to be the top priority for any of the participants.

As robustness check, we also investigate a hybrid model that combines own payoff with an other-regarding preference u_i^m and estimate regressions of the form:

$$\Pr(y_i = 1) = \alpha_0 + \alpha_1 \Delta\pi_i + \alpha_2 (\Delta\pi_i)^2 + \beta_1 \Delta u_i^m + \beta_2 (\Delta u_i^m)^2 + \beta_3 \Delta\pi_i \Delta u_i^m + \varepsilon_i \quad (6)$$

where $\Delta\pi_i$ and Δu_i^m represent the material payoff and utility gain associated with choice $y_i = 1$, respectively. As before, all Δu_i^m are normalized to have the same unit standard deviation. Model (6) is estimated with preference archetypes (2) to (6) above as well as with altruistic preferences and with inequality aversion.¹⁰

For archetypes, the results, not shown here to save space, confirm earlier findings: efficiency considerations are significant in T2 and especially T1; invidious preferences are significant in treatment T3; maximin and equity-only preferences appear with the wrong sign, especially in treatment T1, a finding that is consistent with spiteful preferences; and there is little evidence of homophily. Results regarding altruistic preferences are contrasted. In treatment T3 altruism has the wrong sign, a result consistent with spiteful preferences. But in treatments T2 and especially T1, Δu_i^m becomes positively significant while $\Delta\pi_i$ no longer is. This suggests that altruist preferences do a reasonably good job of predicting participants' choices in T2 and T1 – but not in T3. In contrast, inequality aversion Δu_i^m is either statistically non-significant, or appears with the wrong sign. How

¹⁰For altruist preferences, we set $u_i \equiv \ln(\pi_i) + 0.5 \ln(\pi_{-i})$ where $\pi_{-i} \equiv \sum_{j \neq i, j \in N_i} \pi_j$. Parameter value 0.5 captures an intermediate level of altruism. For inequality aversion we follow Okada and Reidl (2005) and posit a utility function of the form

$$u_i^{10} \equiv \ln(\pi_i) - \frac{1}{2} \sum_{j \neq i, j \in N_i} |\ln(\pi_i) - \ln(\pi_j)|^- - \frac{1}{3} \sum_{j \neq i, j \in N_i} |\ln(\pi_i) - \ln(\pi_j)|^+$$

and vary N_i to include either only i 's partner or the three individuals affected by i 's choice.

much weight we should ascribe to the latter findings is unclear, given that the experiment was not designed to test altruism or inequality aversion directly, and so has little power.¹¹

4 Discussion and conclusions

We have reported on an experiment designed to test whether people exhibit different other-regarding preferences depending on whether the choices they make are framed as an allocation problem or a partner selection problem. The motivation for this experiment is Bowles' (1998) observation that behavior is less other-regarding in experimental situations approximating a competitive market with many anonymous buyers and sellers.

We find that when choices are framed as a partner selection problem instead of a pure allocation problem, agents are less likely to sacrifice their own material well-being to increase the well-being of others – but more willing to sacrifice a higher payoff to reduce the difference between their payoff and that of others. These findings are broadly consistent with the literature on inequality aversion (e.g., Fehr and Schmidt 1999, Okada and Reidl 2005), but with a twist. Experimental subjects with a higher than average payoff exhibit some altruism or concern for efficiency, but more so in treatments are couched as an allocation problem or a choice of partner in a small group. When asked to select a partner in a large anonymous setting, high payoff players no longer display any sign of altruism and simply maximize their own material payoff. In contrast, subjects with the lowest average payoff display no altruism or concern for efficiency in the small group setting, but exhibit spiteful preferences in a large anonymous setting. In other words, we get the two ‘sides’ of inequality aversion (altruism and spite), but not in the same setting.

These findings raise a number of issues. Fafchamps (2012) argues that economic development requires a change in allocation processes away from allocation within the household or extended family to allocation within markets or within hierarchical organizations. This transformation requires a change in social norms from risk sharing in long-term gift exchange to contract compliance

¹¹The coefficient of correlation between $\Delta\pi_i$ and Δu_i^m exceeds 0.9 for altruist preferences and each inequality aversion measure we tried.

in an anonymous market setting. In a gift exchange allocation process, efficiency requires that individuals make choices that are altruistic or efficiency-seeking. In market exchange, efficiency can be achieved through competition alone; altruistic or efficiency-seeking behavior is not required. To the extent that the behavior of our experimental subjects can be interpreted as reflecting context-specific norms, they fit this pattern to a large extent. We do, however, also find that less fortunate participants occasionally select a partner so as to prevent them from achieving a higher payoff. It is unclear whether competition is sufficient to counter the inefficiency produced by such choices. More research is needed.

The findings also raise a more fundamental question: Why, even with minimal contextualization, do human subjects respond the way they do to differences in the frame in which choices are made? Of particular relevance is the contrast between treatments T2 and T3 which are formally similar but elicit quite different responses from participants. Is it possible that the human brain processes moral choices in a way that systematically reduces altruism and reinforces spite in an anonymous partner-selection setting relative to a small group setting?

The literature on trolley experiments (e.g., Greene 2012) suggests one possible avenue of enquiry so far ignored by economists, namely that people feel less guilty about the consequences of their actions when these consequences seemingly depend on mechanical devices, random events, and choices made by others.¹² If correct, this interpretation suggests that markets – and partner selection problems in large populations – blunt other-regarding preferences by diluting the perceived effect that actions have on the welfare of others, thereby eliciting less guilt for failing to follow norms of acceptable behavior that apply in small groups. We offer in Appendix 2 a simple model of such preferences.¹³ Further work is needed on the origin of other-regarding preferences, and especially

¹²Mikhael (2011) goes so far as to suggest that this is because the human brain processes moral choices of cause and effect by applying syntactic rules. This means that ‘pushing the man to his death’ generates more guilt than ‘pushing the button that activates the lever that opens the door that pushes the man to his death’ even though the ultimate consequence is the same.

¹³The difference between treatments T2 and 3 similarly is in the *process* by which payoff are generated, not in the actual payoffs themselves. Dilution as defined here can thus be seen as a situation in which people’s preferences depend not just on payoffs but also on the way these payoffs are obtained. If this interpretation is correct, this paper can be seen as following in the footsteps of Charness and Rabin (2003) who demonstrated that people have preferences over process, not just final outcomes.

the extent to which they are shaped by the decision environments over which altruistic norms apply.

5 References

References

- [1] Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström, (2008). "Eliciting Risk and Time Preferences," *Econometrica*, 76(3): 583-618, 05.
- [2] Becker, G. (1973), "A Theory of Marriage: Part I", *Journal of Political Economy*, 81(4): 813-46.
- [3] Blanchflower, D.G. and A.J. Oswald (2004), "Well-Being over Time in Britain and the USA", *Journal of Public Economics*, 88(7-8), July 2004
- [4] Bolton, G.E. and A. Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, 90(1): 166-193, March.
- [5] Bowles, S. (1998). "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions", *Journal of Economic Literature*, 36(1): 75-111.
- [6] Charness, G. and M. Rabin (2002). "Understanding Social Preferences With Simple Tests", *Quarterly Journal of Economics*, 117(3): 817-869,
- [7] Camerer, C., Thaler, R. H. (1995). "Anomalies: Ultimatums, dictators, and manners", *Journal of Economic Perspectives*, 9(2): 209-219.
- [8] Camerer, C. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton NJ.
- [9] Fafchamps, M. and F. Shilpi (2008). "Subjective Welfare, Isolation, and Relative Consumption", *Journal of Development Economics*, 86(1): 43-60, April
- [10] Fehr, E. and K.M. Schmidt (1999). "A theory of fairness, competition, and cooperation", *Quarterly Journal of Economics*, 114: 817-868.

- [11] Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. Smith, Hill, K. Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J.Q., Tracer, D. (2005), "Economic man in cross-cultural perspective: Behavioral experiments in 15 small-scale societies", *Behavioral and Brain Sciences*, 28(6): 795-815.
- [12] Hoffman, E., K. McCabe and V. Smith (1996), "Social distance and other-regarding behavior in dictator games", *American Economic Review*, 86: 653-60
- [13] Greene, Joshua (2010). "Of Trolleys and Cheaters: Automatic and Controlled Influence on Moral Judgment", video available at <http://www.santafe.edu/research/videos/play/?id=a768d1f4-06e5-4533-8ebe-9ba3d3a5c131>,
- [14] Greene, Joshua (2012). *For The Greater Good: How the Moral Brain Works and How it Can Work Better*, Penguin Press, New York, (forthcoming)
- [15] Ledyard, John O. (1995). "Public Goods: A Survey of Experimental Research", in J. Kagel and A. E. Roth, *Handbook of Experimental Economics*, 111-194. Princeton: Princeton University Press.
- [16] Luce, R.D., (1959). *Individual choice behavior*, John Wiley., Oxford, England
- [17] Luce, R. Duncan (2010). "Behavioral Assumptions for a Class of Utility Theories: A Program of Experiments", *Journal of Risk and Uncertainty*, 41: 19-27.
- [18] McPherson, M., L. Smith-Lovin and J. M. Cook (2001). "Birds of a Feather: Homophily in Social Networks", *Annual Review of Sociology*, 27: 415-44.
- [19] Mikhail, John (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*, Cambridge University Press, Cambridge
- [20] Null, Claire (2012). "Warm glow, information, and inefficient charitable giving", *Journal of Public Economics*, (forthcoming)

- [21] Okada, Akira and Riedl, Arno, (2005). "Inefficiency and social exclusion in a coalition formation game: experimental evidence," *Games and Economic Behavior*, 50(2): 278-311, February.
- [22] Smith, Adam. (1759) 1976. *The theory of moral sentiments*, Liberty classics, edited by D. D. Raphael and A. L. Mactie. Indianapolis:Liberty Press.
- [23] Smith, Adam. (1776) 1909. *Wealth of nations*, Harvard Classics. volume 10, edited by C. J. Bullock. New York: P. F Collier
- [24] Smith,V.L. (1962). "An experimental study of competitive market behavior", *Journal of Political Economy*, 70(2): 111-37.
- [25] Smith, V. (1998), "The Two Faces of Adam Smith", *Southern Economic Journal*, 65(1): 1-19
- [26] Young, P. and M.A. Burke (2001), "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture", *American Economic Review*, 91(3): 559-573

6 Appendix 1. Experimental protocol

We consider three different versions of the second stage. The first is the "*partnership formation in large groups*" version, whereby participants are organized in groups of 24 divided equally into A and M categories. participants are asked to choose a partner type from the other category . If their choice is implemented, we randomly match them with one of the 12 people from the other category. The second version is the "*partnership formation in small groups*" version, whereby participants are put in groups of 4 (a high and a low type in the A category and a high and a low type in the M category). participants are asked to choose whether they would like to form a partnership with the high or low type in the other category. The last version is the "*earnings division in small groups*" version whereby participants are not asked to choose a partner, but are simply asked to choose between two divisions of earnings between four players. The scenarios we propose in these three treatments are perfectly equivalent in terms of their implications for efficiency and income distribution. The only variation is the salience of these implications. Importantly, we provide the same information in all treatments.

Treatment T3 - Partnership formation - large groups

Here each participant is asked to report whether they would prefer forming a partnership with a low or high type from the other category. The scenarios are presented on an answer sheet (see appendix 1). They are shown the distribution of earnings associated with the four different types of partnership (high A - high M , low A - low M , high A - low M , and low A - high M), including those that do not involve them. These choices are illustrated graphically with coloured pies. The size of the pie represents the total earnings to be shared. Each earnings division corresponding to each partnership is represented with a pie division. Participants are asked to tick one of two boxes (top 50% or bottom 50%) at the bottom of the answer sheet.

The partnership formation is implemented as follows. One of the categories (A or M) is chosen randomly ex-post to be the leader in partnership formation (which means that the partnerships will be implemented according to their reported preferences, irrespectively of the reported preferences

made by participants in the other category). If there is more demand for one type than is available, then the scarce type is allocated in a random manner between those who have expressed a preference for it.

Table 2 describes the different parameter configurations corresponding to the different situations. Note that the payoffs are not always symmetric for low types. To be able to distinguish between selfish preferences and maximin payoff, we designed a situation where the payoff of the low type individual is not the minimum payoff. In those situations, the gains from partnership are not shared equally between the A and M partners. The addition partner gets a larger share. In those cases, a low A type who wishes to maximise the minimum payoff would choose a high M type, while if she wishes to maximise her own payoff, she would choose a low M type.

Treatment T2 - Partnership formation - small groups

In the second treatment, participants are told that they are randomly allocated in groups of 4, composed of 2 people who did the multiplication task (one high, one low) and 2 people who did the addition task (one high, one low). They are asked again whether they would prefer forming a partnership with a low or high type from the other category. The scenarios are presented on an answer sheet (see appendix 1) in a manner similar to treatment 1, except that we now write explicitly the implication of the partnership decision for the other people in the group. They are asked to tick one of two boxes (top 50% or bottom 50%) at the bottom of the answer sheet (see Appendix 1)

Treatment T1 - Earnings division

In treatment participants are told that they are randomly allocated in groups of 4, composed of 2 M participants (one high and one low) and 2 A participants (one high and one low). They are asked to choose between two distributions of earnings that correspond to the earnings distribution in T3 and T2 (see Appendix 1). Each earnings distribution is represented by a pie division. The main difference with T3 and T2 is that to each choice is associated a single pie that represents the division of earnings between the *four* people in the group, instead of two pies for each of the two partnerships. Thus, the implication of the decision for efficiency and income distribution is most

salient in this last treatment. This treatment is also the closest to the dictator game designs used in the literature on social preferences (references...)

7 Appendix 2. Dilution

In this appendix we illustrate how altruism can become diluted in an anonymous partner-selection setting.¹⁴ The trolley experiments (e.g., Greene 2012, Mikhael 2011) suggest that people feel less guilt when the consequences of their action involve mechanical devices, random events, or choices made by others. This suggests that, say, pushing an anonymous person to their death generates more guilt than pushing a button that randomly selects one of M anonymous persons to be pushed to their death. A similar contrast characterizes the difference between treatments T2 and T3: selecting in T2 a partner that leaves other experimental subjects a lower payoff may generate more guilt than indicating in T3 a preference for a partner type that, de facto, removes one partnership from the choice set of others and has similar consequences on randomly selected individuals.

To illustrate how this idea can be formalized, we construct preferences in which individuals value the consequence of their action on others differently depending on whether they affect, with certainty, one person or one of M randomly selected individuals. Let $W_2(h)$ denote the utility gain from choosing a ‘high’ partner h instead of a ‘low’ partner l . Consider treatment T2 and let this choice be efficient, so that for a subject with sufficiently altruistic preferences, we have:

$$W_2(h) \equiv \pi_i^h - \pi_i^l + \beta \sum_{j \neq i, j \in N_i} (\pi_j^{h_i} - \pi_j^{l_i}) > 0$$

where $\beta \leq 1$ is a parameter capturing the strength of altruism, N_i is the set of four subjects that includes i , and $\pi_j^{h_i} - \pi_j^{l_i}$ is the effect that player i has on the payoff of player j in N_i when choosing h .

In treatment T3, the effect on the payoff of others is essentially the same as in treatment T2 but i does not know which exact players will be affected. The expected efficiency gain from choosing a

¹⁴More sophisticated models can be written – e.g., models in which subjects have preferences on whether they interfere or not with other subjects’ choices – but they would take us too far from the object of this paper, which is primarily empirical.

partner h can now be written:

$$W_3(h) \equiv \pi_i^h - \pi_i^l + \beta \sum_{j \neq i, j=1}^{24} E[\pi_j^{h_i} - \pi_j^{l_i}]$$

Realized payoffs are as in treatment T2, but the identity of the three individuals affected by i 's decision has not yet been determined. It is this difference that opens the door to a possible dilution effect as follows.

Let us first consider i 's randomly assigned partner and, without loss of generality, assume that this person belongs to category M . The effect of i 's choice on this person's payoff is $\pi_j^{h_i} - \pi_j^{l_i}$; other possible high M partners are unaffected by i 's decision so that for them the effect is 0. The total effect of i 's choice on the expected payoff of high M subjects is thus:

$$\sum_{j=1}^6 E[\pi_j^{h_i} - \pi_j^{l_i}] = \sum_{j=1}^6 \frac{1}{6} (\pi_j^{h_i} - \pi_j^{l_i}) = \pi_j^{h_i} - \pi_j^{l_i} \quad (7)$$

Similar calculations can be done for subjects in the other two category \times type groups. We get:

$$W_2(h) = W_3(h)$$

which predicts that, in the absence of dilution, individuals should make identical decisions under treatments 2 and 3.

Let us now introduce a dilution parameter $\alpha \geq 1$ and rewrite (7) as:

$$\sum_{j=1}^6 E^\alpha[\pi_j^{h_i} - \pi_j^{l_i}] = \sum_{j=1}^6 \left(\frac{1}{6}\right)^\alpha (\pi_j^{h_i} - \pi_j^{l_i}) \leq \pi_j^{h_i} - \pi_j^{l_i} \quad (8)$$

with strict inequality if $\alpha > 1$. The effect on the other two category \times type groups can be handled in the same way. Equation (8) is equivalent to positing that individuals underweigh the probability that they affect other players, and is formally similar to probability weighting in prospect theory.

With α large enough, $\sum_{j=1}^6 \left(\frac{1}{6}\right)^\alpha (\pi_j^{h_i} - \pi_j^{l_i})$ tends to 0 and we have:

$$W_3(h) = \pi_i^h - \pi_i^l$$

which corresponds to selfish preferences: with enough dilution, people no longer take into account any efficiency cost they impose on others, and pursue their own material welfare only.

8 Reject

8.0.1 Hybrid preferences

So far we have focused on the simple preference archetypes used in designing the experiment. As robustness check, we now examine the extent to which experimental results can be explained by alternative other-regarding utility functions, namely, altruistic and inequality averse.¹⁵

For altruistic preferences, we consider a utility function of the form $u_i^7 \equiv \ln(\pi_i) + 0.5 \ln(\pi_{-i})$ where π_{-i} is the average payoff of the three other participants affected by i 's choice, i.e., $\pi_{-i} \equiv \sum_{j \neq i, j \in N_i} \pi_j$. Value 0.5 is selected to capture an intermediate level of altruism.¹⁶

We consider several versions of inequality aversion. The first version posits that player i only cares about the payoff of his/her selected partner k . We follow Okada and Reidl (2005) and posit a utility function of the following form:

$$u_i^8 \equiv \pi_i - \frac{1}{2} |\pi_i - \pi_k|^- - \frac{1}{3} |\pi_i - \pi_k|^+ \quad (9)$$

where $|\pi_i - \pi_k|^-$ is $\pi_k - \pi_i$ if $\pi_i < \pi_k$ and 0 otherwise, and $|\pi_i - \pi_k|^+$ is $\pi_i - \pi_k$ if $\pi_i > \pi_k$ and 0 otherwise. Parameter values 1/2 and 1/3 are selected to capture an intermediate level of inequality aversion. The second version u_i^9 replaces π by $\ln(\pi)$ throughout in (9). The third version includes all three participants that i may affect by his/her choice. This yields:

$$u_i^{10} \equiv \ln(\pi_i) - \frac{1}{2} \sum_{j \neq i, j \in N_i} |\ln(\pi_i) - \ln(\pi_j)|^- - \frac{1}{3} \sum_{j \neq i, j \in N_i} |\ln(\pi_i) - \ln(\pi_j)|^+$$

For each utility function considered in this paper – the six archetypes plus the above four utility functions – we calculate $\Delta u_i \equiv u_i^h - u_i^l$, that is, the utility gain from choosing a high type as partner (or equivalent in T1). We then estimate regressions of the form:

$$\Pr(y_i = 1) = \alpha_0 + \alpha_1 \Delta \pi_i + \alpha_2 (\Delta \pi_i)^2 + \beta_1 \Delta u_i^m + \beta_2 (\Delta u_i^m)^2 + \beta_3 \Delta \pi_i \Delta u_i^m + \varepsilon_i \quad (10)$$

for $m = \{2, \dots, 10\}$ separately for each of the three treatments. As before, all Δu_i^m are normalized to have the same unit standard deviation. Own payoff is included in all regressions since it has been

¹⁵Because the experiment was not designed to test these preferences, power may be weak, especially for preference functions that happen to be correlated with own payoffs in our experiment.

¹⁶Other values of the altruism coefficient do not affect the conclusions.

shown to be dominant in all treatments and payoff categories. Regression model (10) is best seen as a hybrid model allowing different types of preferences to influence choices. Since the regression coefficients themselves are difficult to interpret, we present instead the average marginal effect of $\Delta\pi_i$ and Δu_i^m on $E[y_i]$.

We also present the marginal rate of transformation (MRT) between the two (selfish and other-regarding) preferences $\frac{d\Delta u_i^m}{d\Delta\pi_i} = -\frac{\partial E[y_i]/\partial\Delta\pi_i}{\partial E[y_i]/\partial\Delta u_i^m}$. The MRT measures by how much Δu_i^m has to increase to keep $E[y_i]$ constant in order compensate for a unit decrease in own payoff $\Delta\pi_i$. It is the slope of the indifference curve between Δu_i^m and $\Delta\pi_i$ evaluated at the average.¹⁷ The more negative the MRT is, the more Δu_i^m has to increase, and thus the less i cares about Δu_i^m . If the slope is -1 , participants care on average equally for their own payoff and u_i^m .

Results are given in Table 9. The first five sets of marginal effects correspond to the five archetypes used above. Earlier findings are confirmed: efficiency considerations are significant in treatments T2 and especially T1 (MRT close to -1); invidious preferences are significant in treatment T3; maximin and equity only preferences appear with the wrong sign, especially in treatment T1, a finding consistent with invidious preferences; and little evidence of homophily.

Results using an altruist utility function are contrasted. In treatment T3 altruism has the wrong sign, a result consistent with invidious preferences. But in treatments T2 and especially T1, Δu_i^m becomes positively significant while $\Delta\pi_i$ no longer is. This suggests that the altruist utility function does a good job of predicting participants' choices in T2 and T1 – but not in T3. This is remarkable given that the correlation coefficient between Δu_i^m and $\Delta\pi_i$ is 0.9 for altruism.

We are less successful with inequality, which is the focus of the last three results of Table 9. Inequality aversion with respect to the chosen partner gets a small coefficient and is never statistically significant. If we use logs instead of levels, the estimated marginal effect becomes significantly negative. The same holds if we compute inequality aversion with respect to the three participants affected by i 's choice: the marginal effect is significant but with the wrong sign. How much weight we should ascribe to these findings is unclear, given that the coefficient of correlation

¹⁷Since positive values of the MRT are meaningless, we only report negative values.

between each of the three inequality aversion measures and own payoffs exceeds 0.9. The experiment was not designed to test inequality aversion directly, and so has little power.¹⁸

¹⁸We also estimate regression (10) separately for each of the three subject types. Marginal effects and MRT's are summarized in Table A1 in Appendix. Although there is some large swings in estimated marginal effects, the same pattern repeats itself: more evidence of efficiency considerations in T2 and T1, more evidence of invidious preferences in T3. We also find that high types are more concerned about inequality in treatment T3 than types with low average payoffs. Altruist preferences are strongly rejected for all types in T3, i.e., the marginal effect is significantly negative and large. Altruist preferences are significant and positive for some types and treatments, but there appears to be some parameter instability across regressions, probably due to high multicollinearity between Δu_i^7 and $\Delta \pi_i$.