UNIVERSITY *of York*

Discussion Papers in Economics

No. 18/11

Giving to Varying Numbers of Others

Matthew Robson
John Bone

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Giving to Varying Numbers of Others

Matthew Robson and John Bone[‡]

Working Paper
October 3, 2018

Abstract

Within a modified N person dictator game, we test the extent to which giving behaviour changes as the number of recipients varies. Using a *within-subject* design, in an incentivised laboratory experiment, *individual-level* preference parameters are estimated within *five* alternative utility functions. Both *goodness-of-fit* and *predictive accuracy* of each model are analysed, with the 'best' model identified for each individual. The Dirichlet distribution is proposed as a random behavioural model to rationalise *noise*; with parameters accounting for differential error arising from the *complexity* of decision problems. Results show that, on average, participants are willing to give more *total* payoffs to others as the number of players increase, but not maintain *average* payoffs to others. Extensive heterogeneity is found in individual preferences, with no model 'best' fitting all individuals.

Keywords: Distributional Preferences, Prosocial Behaviour, Group Size, Experimental Economics, Altruism, Social Welfare Function.

JEL Classification: C72, C91, D63, D64, I31.

[†]Corresponding Author: Matthew Robson, Department of Economics and Related Studies, University of York. Email: matthew.robson@york.ac.uk. Website: www.mrobson92.com.

[‡]We would like to thank Richard Cookson and John Hey for offering invaluable comments and guidance. Thanks also to Andrew Jones, Peter Moffatt and Erik Sørensen, alongside seminar participants at the Young Economists' Meeting, Masaryk University; SSSI, University of Bonn; Nordic Conference on Behavioural and Experimental Economics, University of Southern Denmark; and FUR, University of York for their helpful comments and enlightening discussions. For the financial support from the ESRC White Rose DTC (ES/J500215/1) and the Department of Economics and Related Studies, University of York, we also give our thanks. All errors remain our own.

1 Introduction

Often our behaviour has consequences for the happiness or misery of others in society. In those instances our behaviour is shaped not only by our preferences, but by the number of others whom we affect. This is particularly the case in the context of prosocial behaviour. In giving to a particular individual we forego potential payoffs to ourselves and others. As the number of others increase, so to does the complexity of the decisions we have to make. Not only must we consider the trade-offs we are willing to make between our self and others, but between-others. Are we willing to let the average amount to others decrease, in order to maintain the same amount for ourselves, or is there a minimum acceptable level we must give to all others?

This paper seeks to model prosocial behaviour as the number of recipients of giving increases. Individual-level preference parameters are estimated within five alternate CES utility functions. Preferences accounting for *inequality aversion*, the trade-off between equality and efficiency, and *self-interest*, the weight on the self in relation to others, are central to each functional form. However, additional preference parameters are incorporated within *extended* models, to account for alternative behavioural responses to changes in N. The first models the distinction between *self-other* and *between-other* inequality aversion. The second incorporations *congestion*, the trade-off between *average* and *total* payoffs to others. The third accounts for *minimum threshold* levels of giving, denoting absolute levels of payoffs which are deemed necessary to distribute to each player.

The relative goodness-of-fit and predictive power of each model is tested, allowing for the identification of 'types' of individuals. This approach allows the flexibility to explain heterogeneity in individual behaviour not only through preference parameters within a particular assumed model, but between different behaviour assumptions made in alternative models. To account for *noise* in decision making, the Dirichlet Distribution is formulated as a random behavioural model. Additional error parameters are incorporated, which allow for differential error as the *complexity* of decision making increases.

To observe prosocial behaviour an incentivised laboratory experiment is run, in the form of a modified *N-person* dictator game. The *within-subject* design of the experiment varies the *number of players*, over 45 rounds of decision problems. Two treatments are conducted, the *multiple slider* and *single slider* treatments. The former allows for complex between-other distributional decisions to be made, in addition to self-other decisions, for 2, 3 and 4 player games. The latter simplifies the decision problem, but allows for an increased variation in the number of players: 2, 3, 4, 6 and 12. The experimental design specifically allows for the testing of both goodness-of-fit and predictive accuracy of the alternative utility functions to be compared.

Within the literature surrounding *dictator games*, the number of players within the experiment is often held constant.[1] Some papers have, however, varied the number of players within the experiment. Charness and Rabin (2002) run a host of "simple" experimental games, within which are two and three person dictator games. Fisman, Kariv, and Markovits (2007) investigate modified two and three person dictator games, using budget sets, while Macro and Weesie (2016) use batteries of pairwise questions for two-player and four-player dictator games. The increase in the number of recipients allows for the identification of prosocial behaviour relating to *between other* trade-offs, alongside the usual *self-other* trade-offs. Panchanathan, Frankenhuis, and Silk (2013) increase the number of dictators, and find that dictators transfer less when there are more dictators, while Cason and Mui (1997) run both team and individual dictator games. Schumacher et al. (2017) motivate well an experiment where a 'decider' chooses the provision of a good between themselves and a 'receiver', where such provision comes at a cost to a group of 'payers'. They utilize a general form of the Andreoni and Miller (2002) utility function, and identify a substantial fraction of subjects which are "insensitive to group size", through a between-subject treatment design which varies the number of 'payers'.

The effect of changes in *group size* on behaviour has also been investigated in parallel literatures. Papers by Isaac and Walker (1988) and Isaac, Walker, and Williams (1994), amongst others, identify group size effects in the context of public goods games. N-person prisoner's dilemmas are studied by many, including Marwell and Schmitt (1972) and Bonacich et al. (1976). The experimental oligopolies literature identifies the effects of group size on cooperation, such as Fouraker and Siegel (1963) and Dolbear et al. (1968). The size of the group, clearly plays an important role in the decision making process. It is, therefore, an integral component within models which strive to explain behaviour.

Andreoni (2007) addresses this observation, proposing a CES utility function to explain prosocial behaviour as the number of recipients of giving increases. A modified N-person dictator game was run, where participants chose to *hold* a number of tokens (from a set budget), *passing* the remainder to a group of other players. The budget sets, prices of giving and number of other participants varied through the 24 rounds. Preferences were estimated within the utility function proposed, which incorporated a *congestion* parameter, b, signifying the extent to which the *total* or *average* payoffs to others were considered.

While the model proposed in Andreoni (2007) allows for extensive heterogeneity amongst individuals, alternative models could better explain the behaviour of particular individuals. Preference

[1] Two player dictator games are frequently used to identify prosocial behaviour, for example: Forsythe et al. (1994), Hoffman et al. (1994) and Andreoni and Miller (2002). Extensions of such dictator games to incorporate multiple players (for a review see Engelmann and Strobel (2007)) include: Engelmann and Strobel (2004), Fehr, Naef, and Schmidt (2006) and Karni, Salmon, and Sopher (2008). Erkal, Gangadharan, and Nikiforakis (2011) and Barr et al. (2015) also utilise four-person dictator games, where entitlements are earned.

parameters can be estimated within alternative functions and the relative goodness-of-fit of each model tested at an individual level. Hey and Orme (1995) compare the goodness-of-fit of five alternative models in the context of risk. Hey and Pace (2014) conduct similar work, focusing upon ambiguity, but highlight the importance of considering both the goodness-of-fit and predictive power of each model. By comparing alternative models the 'best' model can be identified for each individual; enabling the observation of 'types' of individual based on the differing behavioural assumptions made.

This research seeks to contribute to the above literature, by intertwining important considerations of the papers above. We propose a within-subject design which allows for the complexities of *self-other* and *between-other* trade-offs to be observed as the number of recipients increase. Individual-level preference parameters are estimated within five alternative models, allowing for both the goodness-of-fit and predictive power of the respective models to be analysed. In addition, the Dirichlet Distribution is formulated to account for *noise* in decision making, incorporating parameters which model stochastic responses to the *complexity* of decision problems.

2 Experiment

The general form of the experiment is a modified N-person 'dictator' game; where individuals are required to make distributional decisions amongst participants within a group. 'Dictators' are given a budget, m, which they must distribute amongst N players; themselves and n others, the 'recipients'. Dictators choose *allocations*, x_i, for each player in the group; where $i \in [1, ..., N]$ and $\sum_i^N x_i = m$. These *allocations* are then divided by the corresponding *divider*, $1/\pi_i$, to give the *payoff*, $\pi_i x_i$, to each Player i.[2] It is the *multipliers* (π_i) which make the dictator game 'modified', as through them the *relative prices of giving* to each player can vary; meaning that equality-efficiency trade-offs need to be made by participants.

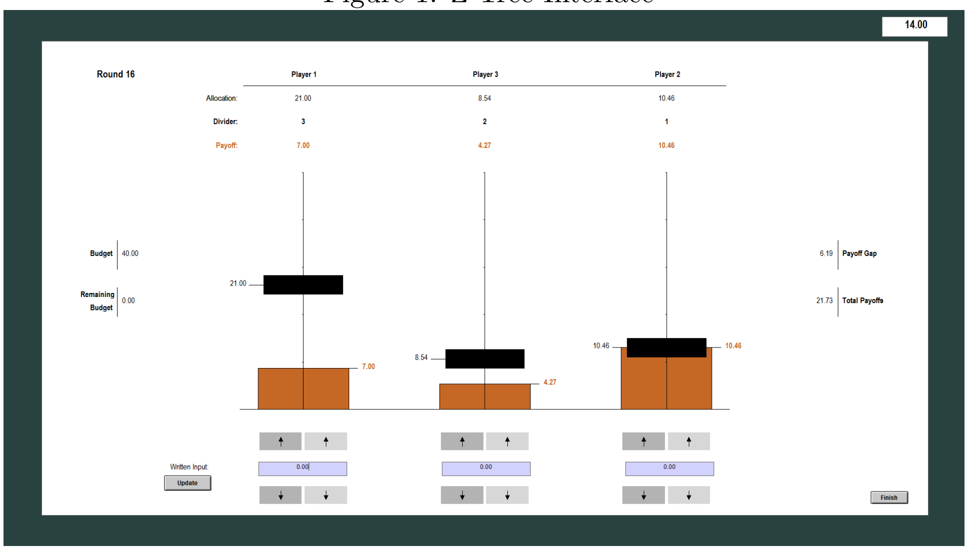
There are two within-subject treatments, across 45 rounds of the experiment; the *multiple slider* treatment (30 rounds) and *single slider* treatment (15 rounds). In each round the participants are randomly assigned to groups, made up of N participants. Between rounds the dividers, $1/\pi_i$, change, ensuring the relative price of giving to each player varies. The budget, m, also changes, varying the average (per player) and total amounts available to distribute.

Participants made their decisions on a computerized Z-Tree interface. They were given extensive paper instructions (found in Appendix A.1), followed by an interactive on-screen tutorial to enable them to use the interface. A screenshot of the interface, from the *multiple slider* treatment, is shown in Figure 1. In this example, there are three players, Player 1, 2 and 3, amongst whom the

[2] The reason that *dividers* ($1/\pi_i$) are used, rather than *multipliers* (π_i) are due to visual constraints on the Z-Tree interface. Multipliers are, however, used throughout the theory, for notational ease.

'dictator' must make allocations, so that the remaining budget reaches zero. Each player has a divider (changing every round), which is used to calculate the payoff to that player. Allocations can be made by using: the slider, arrow keys and written input. The slider (the black bar) can be dragged to make allocations, the arrow keys clicked to make incremental changes (0.01 or 0.1), and the written input used to type exact amounts. The *single slider* treatment differs in that there is only one slider, written input and set of arrow keys; that determines the allocations (and hence payoffs) to the *self*. The remaining budget is then split equally between the *recipients*. Calculations of the payoffs are made automatically, and are shown by both the orange numbers and by the height of orange bars. The payoff gap, the highest payoff minus the lowest payoff, and the total payoffs, the sum of all payoffs, are shown. All allocations, payoffs and budgets are shown in pounds and pence.

Figure 1: Z-Tree Interface



Within each of the seven experimental sessions there were twelve participants. Each participant made individual decisions; *as if* they were the 'dictator'. One individual's decisions, from each group, was randomly selected (at the end of the experiment) to determine the payoffs of each member of their group. Then *one round* was randomly selected to determine the 'dictators', and their distributional decisions determined the payoffs of the participants in their group. In this way, each distributional decision participants made had an equal chance of determining their payoff and the recipients payoffs, and hence they were fully *incentivised*. Players within each group were randomly matched each round. Each decision made was entirely anonymous and without feedback; participants neither knew the decisions of any other participants nor the identity of the 'dictator' in any round. Removing considerations of reputation and reciprocity, allowing for the identification of 'pure' altruism.

The experiments were run in the EXEC laboratory at the University of York. Randomised invites were sent out, using hroot (Hamburg Registration and Organisation Online Tool), amongst a pool of 2,692 users. Seven experimental sessions were run between the 28th of March and the 6th of April 2017, with twelve participants in each session, to reach a sample size of 84 participants.[3] Each session required twelve participants in order to run. Due to a lack of participants one session had to be cancelled. A further 30 users were invited, as reserve participants. Nine users who showed up could not take part in the experiment, so they (and the six in the cancelled experiment) received show-up fees. The average payoff per participant was £15.45. Details of the demographic characteristics of the sample can be found in Appendix A.2.

2.1 Multiple Slider Treatment

Table 1 shows how the design parameters change throughout the 30 rounds of the *multiple slider* treatment. The number of players, N, changes every ten rounds. The change in the budget, m, is shown alongside the change in the budget per player, m/N. The dividers, $1/\pi_i$, for each player (1 to 4) are shown, alongside the *relative cost* (of giving). The *relative cost*, p, shows the cost in payoffs to the 'self' of increasing the payoffs to each of the 'others', where $p = \pi_1(\sum_{j=2}^{N} 1/\pi_j)$. The *average relative cost*, p/n, shows this cost per 'recipient'.

Particular variations in the design parameters is ensured, to enable better identification of preferences parameters, in the models shown in Section 3. Between differing N, both overlap and variation is ensured in the budget variables, dividers and relative costs. This ensures that equality-efficiency, self-other and between-other trade-offs need to be made, alongside considerations of average vs total payoffs to others and minimum levels of payoffs to be given to recipients. In particular the design allows for some design parameters to remain identical between problems with differing N, while varying others. The sets (of rounds) [1,20,30], [3, 16, 23], [3, 16, 23] and [2, 22] remain constant in m and p/n. Sets [10, 13], [11, 25] and [2, 15] are constant in m and p. Both m/N and p/n remain constant in [17, 26] and [1, 11, 21], while m/N and p is constant in [6, 27].

Rounds in each set of N maintained the order shown in Table 1, to ensure that the grouping procedure was feasible and transparent to participants. The order of N was, however, randomised between experimental sessions, to enable the testing of order effects on decision making. Further randomisation was applied to the screen order of the players, allowing for the effect of screen position (i.e. left, middle, right) and player name (i.e. 2, 3, 4) to be tested.

[3] One participant had to be dropped from the analysis due to concerns of contamination.

Table 1: Experiment Design Parameters

Round	Players, N	Budget		Dividers, $1/\pi_i$				Relative Cost	
		m	m/N	P1	P2	P3	P4	p	p/n
1	2	30	15	1	1	.	.	1	1
2	2	40	20	1	2	.	.	2	2
3	2	40	20	2	1	.	.	0.5	0.5
4	2	22	11	1	3	.	.	3	3
5	2	22	11	3	1	.	.	0.33	0.33
6	2	70	35	1	4	.	.	4	4
7	2	70	35	4	1	.	.	0.25	0.25
8	2	12	6	1	3	.	.	3	3
9	2	12	6	3	1	.	.	0.33	0.33
10	2	35	17.5	1	1	.	.	1	1
11	3	45	15	1	1	1	.	2	1
12	3	35	11.67	1	2	2	.	4	2
13	3	35	11.67	2	1	1	.	1	0.5
14	3	40	13.33	1	2	3	.	5	2.5
15	3	40	13.33	2	3	1	.	2	1
16	3	40	13.33	3	1	2	.	1	0.5
17	3	105	35	1	2	4	.	6	3
18	3	105	35	2	4	1	.	2.5	1.25
19	3	105	35	4	1	2	.	0.75	0.38
20	3	30	10	1	1	1		2	1
21	4	60	15	1	1	1	1	3	1
22	4	40	10	1	2	2	2	6	2
23	4	40	10	2	1	1	1	1.5	0.5
24	4	45	11.25	1	2	1	2	5	1.67
25	4	45	11.25	2	1	2	1	2	0.67
26	4	140	35	1	2	3	4	9	3
27	4	140	35	2	3	4	1	4.0	1.33
28	4	140	35	3	4	1	2	2.3	0.78
29	4	140	35	4	1	2	3	1.5	0.5
30	4	30	7.5	1	1	1	1	3	1

2.2 Single Slider Treatment

The *single slider* treatment is a simplified version of the above, which increases the variation in the number of players, N. There are 15 rounds within the treatment, where $N \in [2, 3, 4, 6, 12]$ and each N has a set of three rounds. Rather than the more complex decision problem, where the participant must make distributive decisions separately for each player, only one slider (written output and set of arrow keys) is used to make decisions. This slider denotes the share between the self and others, where each of the others will get an equal share of the remainder of the budget, not allocated to the self. As before the budgets and dividers change each round, however, there are only the divider to the self $1/\pi_1$, and divider for each other, $1/\pi_o$, as the dividers are the same for each other player.

Both budget and divider set were randomly generated in each round, for each participant. Each set of three rounds, for each N, consisted of three sets of dividers $[1/\pi_1, 1/\pi_o]$: in the first [1,1], the second [1, A] and the third [B, 1]. The dividers A and B are uniformly and independently drawn from the set [2,3,4]. The budget, m, is similarly uniformly drawn. The set, $\ddot{M}$, from which m is drawn differs between rounds, each $\ddot{M} = \{\ddot{m} - 8N + 2N.i\, , \, i \in \{0, 1, ..., 8\}\}$. The $\ddot{m}$, within the calculation of the set differs between rounds, for each N. For the first rounds of each N, where $[1/\pi_1, 1/\pi_o]$ = [1,1], $\ddot{m} = [24, 26, 48, 72, 144]$, for $N = [2, 3, 4, 6, 12]$, respectively. In both the second and third rounds of each N, $\ddot{m} = [40, 60, 80, 120, 240]$, for $N = [2, 3, 4, 6, 12]$, respectively. This random selection of design parameters is used as the number of rounds is limited by time constraints, but through it the variation allows for aggregated analysis to be undertaken.

The two treatments are run as within-subject treatments to allow both goodness-of-fit and predictive accuracy to be analysed. The data from the *multiple slider* treatment is used to estimate preference parameters, allowing for goodness-of-fit measures to be constructed. The estimated preference parameters are then used to predict behaviour in the *single slider* treatment, where variation in N is increased. The combination of fit and prediction is then analysed for each individual to test between proposed utility models, which are formulated in the following section.

3 Utility Functions

Utility functions are proposed below to model behaviour in the experiment. All are within the family of constant elasticity of substitution (CES) models, incorporate preference parameters associated with prosociality and account for variation in N. Five alternative models are proposed. The first is the *standard* function, derived from Andreoni and Miller (2002), which incorporates preference parameters for inequality aversion, r, and self-interest, α. The three subsequent *extended* utility functions build upon the *standard* function by incorporating addition preference parameters, which account for alternative behaviours. The first of these is derived from Fisman, Kariv, and Markovits (2007), it distinguishes *self-other* inequality aversion, r_1, from *between-other* inequality aversion, r_0. The second accounts for *congestion*, b, and is a generalised form of the model in Andreoni (2007). The third takes the form of a Stone-Geary utility function, which originates from Geary (1950) and Stone (1954), accounting for a *minimum threshold* level, τ. In addition to these models, the *amalgamated* function incorporates each of the above preference parameters into a general model. The models are formally presented below, while graphical analysis in Appendix A.3 illustrates the intuition behind the models.

3.1 Standard

The models assume that decisions are based upon distributing allocations: x_i according to a set of preferences parameters and the multiplication factors: π_i, the reciprocals of the dividers, $1/\pi_i$. It is, then, the payoffs, $x_i\pi_i$, amongst the 'self' ($i = 1$) and 'others' ($i \neq 1$) which determine individual utility. Note that the total number of players, N, is distinct from the number of recipients, n. The standard utility function is as follows:

$$U_1 = \left(\sum_{i=1}^{N} \left(\alpha_i (\pi_i x_i)^{-r} \right) \right)^{-\frac{1}{r}} \tag{1}$$

Inequality Aversion is represented by r, where $-1 \leq r \leq \infty$ and $r \neq 0$. When $r = -1$ preferences reflect 'Utilitarianism', where utility is determined by summing payoffs. As r increases more weight is placed upon the payoff of the worst-off, indicating 'Weighted Prioritarianism' (Parfit, 1997), until $r = \infty$ which represents 'Maximin' preferences, where only increases to the worst-off increase utility (Rawls, 1999). Self-interest is represented by α_1, and $\alpha_j = (1 - \alpha_1)/n$, $\forall j \geq 1$ denotes the weight for each 'other', where $\forall i$ $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{N} \alpha_i = 1$. As $\alpha_1 \to 1$, preferences reflect *egoism*, where utility is purely a function of the payoffs to the self. As α_1 decreases α_j increases, reflecting an increased regard for others. 'Cobb-Douglas' preferences are represented when $r \to 0$; which implies that optimal distributions reflect the proportions set by α. Intuitively, r can be thought of as the trade-off individuals are willing to make between *efficiency* and *equality*, across the entire distribution, while α_1 can be thought of as the extent to which the individual weights *themselves*, in relation to *others*.

Given the above utility function and the budget constraint $m = \sum_{i=1}^{n} x_i$, where m is the budget, the following optimal allocations (which maximise utility) can be obtained, $\forall i$:

$$x_i^* = \frac{m}{1 + \sum_{j \neq i}^{N} \left(\frac{\pi_i}{\pi_j} \left(\frac{\alpha_j \pi_j}{\alpha_i \pi_i} \right)^{\frac{1}{1+r}} \right)} \tag{2}$$

3.2 Extended

Building upon the *standard* model are the three *extended* models. Each model adds an additional behavioural assumption, which effects how optimal allocations change as N increases.

3.2.1 FKM

The first model, derived from Fisman, Kariv, and Markovits (2007) henceforth *FKM*, adds assumptions regarding inequality aversion. Two parameters distinguish between *self-other* inequality aversion, r_1, and *between-other* inequality aversion, r_0. This allows for flexibility in decision making, as different equality-efficiency trade-offs can be made, depending on who the trade-off concerns. For example, an individual may prioritise efficiency between the self and others, but want to ensure equality between others. The model is below:

$$U_{F1} = \left(\alpha_1 (\pi_1 x_1)^{-r_1} + \alpha_0 \sum_{i=2}^{N} \left(\alpha_i' (\pi_i x_i)^{-r_0} \right)^{r_1/r_0} \right)^{-\frac{1}{r_1}} \tag{3}$$

Similar to the *standard* model $-1 \leq r_1, r_0 \leq \infty$ and $r_1, r_0 \neq 0$. As $r_1, r_0 \to -1$ efficiency is prioritised, while when $r_1, r_0 \to \infty$ equality becomes paramount. Self-interest, α_1, is as before, but now α_0 shows the aggregate regard for others $(1 - \alpha_1)$. Individual *between-other* weights are given by α_i', where $\alpha_i' = \upsilon_i / \sum_{j \neq 1}^{N} \upsilon_j, \forall i > 1$; denoting the relative weight given to each other player, the expected case, which is used throughout the analysis, is that $\upsilon_i = 1, \forall i > 1$, meaning $\alpha_i' = 1/n$. When $r_1 = r_0$ and $\alpha_i' = 1/n$, or $N = 2$ the *FKM* and *standard* models are equivalent.

Optimal allocations are as follows:

$$x_1^* = \frac{m}{1 + \sum_{j \neq 1}^{N} \left(\frac{\pi_1}{\pi_j} \left(\frac{\alpha_0 \alpha_j' \pi_j}{\alpha_1 \pi_1} \right)^{\frac{1}{1+r_1}} \left(\left(\sum_{k=2}^{N} \left(\alpha_k' \left(\frac{\alpha_j' \pi_j}{\alpha_k' \pi_k} \right)^{\frac{r_0}{1+r_0}} \right) \right)^{\frac{r_1 - r_0}{r_0(1+r_1)}} \right) \right)} \tag{4}$$

$$x_{j \neq 1}^* = \frac{m}{1 + \frac{\pi_j}{\pi_1} \left(\frac{\alpha_1 \pi_1}{\alpha_0 \alpha_j' \pi_j} \right)^{\frac{1}{1+r_1}} \left(\sum_{k=2}^{N} \left(\alpha_k' \left(\frac{\alpha_j' \pi_j}{\alpha_k' \pi_k} \right)^{\frac{r_0}{1+r_0}} \right) \right)^{\frac{r_0 - r_1}{r_0(1+r_1)}} + \sum_{l \neq 1, j}^{N} \left(\frac{\pi_j}{\pi_l} \left(\frac{\alpha_l' \pi_l}{\alpha_j' \pi_j} \right)^{\frac{1}{1+r_0}} \right)} \tag{5}$$

3.2.2 Andreoni

Second is the *andreoni* model, a generalised form the model in Andreoni (2007). Here, a 'congestion' parameter, b, is incorporated in the model, where $b \in [0, 1]$. The 'congestion' parameter allows for a distinction between considering the *average* or *total* payoffs to others. As an example, if a 'dictator' distributes £5, out of £10, to themselves in subsequent decision problems with $N = [2, 3, 4]$ then the total payoffs to others are [£5, £5, £5] while the average payoffs are [£5, £2.5, £1.67], respectively. If a 'dictator' wanted to maintain the same average payoffs to others, say £2, they would have to alter the payoff to the self to be [£8, £6, £4], meaning the total payoffs to others would be increasing, as [£2, £4, £6]. The inclusion of b allows preference for trade-offs between the self and average ($b = 0$) or total ($b = 1$) payoffs to others to be incorporated, allowing for the above differential behaviour as N increases. The utility function is as follows:

$$U_{A1} = \left(\alpha_1 (\pi_1 x_1)^{-r} + \sum_{i=2}^{N} \left(\alpha_i (n^b \pi_i x_i)^{-r} \right) \right)^{-\frac{1}{r}} \tag{6}$$

The difference between this and the standard model is the inclusion of n^b, which is a multiplier of the payoffs to others. If $b = 0$, the models are equivalent, but as $b \to 1$ the two diverge as N increases. The model is identical to that of Andreoni (2007) when $\pi_i x_i = \pi_j x_j, \forall i, j > 1 \,\&\, j \neq i$, and indeed is as such for the single slider treatment. The optimal allocations are as follows:

$$x_1^* = \frac{m}{1 + \sum_{j \neq 1}^{N} \left(\frac{\pi_1}{\pi_j} n^{-b} \left(\frac{\alpha_j \pi_j}{\alpha_1 \pi_1} n^b \right)^{\frac{1}{1+r}} \right)} \tag{7}$$

$$x_{j \neq 1}^* = \frac{m}{1 + \left(\frac{\pi_j}{\pi_1} n^b \left(\frac{\alpha_1 \pi_1}{\alpha_j \pi_j} n^{-b} \right)^{\frac{1}{1+r}} \right) + \sum_{k \neq 1,j}^{N} \left(\frac{\pi_j}{\pi_k} \left(\frac{\alpha_k \pi_k}{\alpha_j \pi_j} \right)^{\frac{1}{1+r}} \right)} \tag{8}$$

3.2.3 Stone-Geary

Third is the *stone-geary* model, with a more general CES form to that derived in Geary (1950). The function incorporates a *minimum threshold* level, τ. This is a level below which negative (or undefined) utility would be obtained; therefore, ensuring τ is distributed to each participant is paramount. Above τ indifference curves take the form of the *standard* function, indeed if $\tau = 0$ the two are equivalent. Below is the model:

$$U_{SG1} = \left(\sum_{i=1}^{N} \left(\alpha_i (\pi_i x_i - \tau_i)^{-r} \right) \right)^{-\frac{1}{r}} \tag{9}$$

The model is specified with individual τ_i, but in the analysis we assume $\tau_i = \tau, \forall i$, to reduce the number of estimated parameters. The parameter τ thus signifies a minimum threshold of payoffs for all N. The inclusion of τ allows for behaviour which differs from that in the standard model, as the budget, m, and N change. The higher τ is relative to m the more equally the payoffs will be distributed. Those who have a higher level of self-interest (α_1) will take more for themselves, but only after the minimum threshold has been distributed to all players. The optimal allocations, $\forall i$, are as follows:

$$x_i^* = \frac{m + \sum_{j \neq i}^{N} \left(\left(\frac{\tau_i}{\pi_j} \left(\frac{\alpha_j \pi_j}{\alpha_i \pi_i} \right)^{\frac{1}{1+r}} \right) - \frac{\tau_j}{\pi_j} \right)}{1 + \sum_{j \neq i}^{N} \left(\frac{\pi_i}{\pi_j} \left(\frac{\alpha_j \pi_j}{\alpha_i \pi_i} \right)^{\frac{1}{1+r}} \right)} \tag{10}$$

Due to the form of the model, one particular issue emerges. Given that we assume individuals have a 'true' minimum threshold, say τ^*, it is foreseeable that due to budget restrictions, in a particular decision problem, there is not a sufficient budget in order to meet τ^*. In this case utility is undefined (if $|r| < 1$) as $\tau^* > x_i \pi_i, \forall i$. A natural assumption to then make is that if τ^* is greater than the minimum feasible payoff, say $x_i' \pi_i'$, then $\tau = x_i' \pi_i'$. The solution: $x_i' \pi_i' = m / \sum_i^N (1/\pi_i)$, ensures that $x_i \pi_i = x_j \pi_j, \forall i, j$.[4] In each decision problem then $\tau = min(\tau^*, m / \sum_i^N (1/\pi_i))$. It is then the 'true' minimum threshold, τ^*, which is estimated. An alternative solution, using non-negativity constraints, is in Appendix A.4.

3.3 Amalgamated

While the above separately extend the *standard* model to include r_0, b and τ, it is feasible that participant's behaviour can be explained by a combination of those factors. Here the above utility functions are *amalgamated* into a general functional form, which is as follows:

$$U_{1*} = \left(\alpha_1 (\pi_1 x_1 - \tau_1)^{-r_1} + \alpha_0 \sum_{j=2}^{N} \left(\alpha_j' \left(n^b (\pi_j x_j - \tau_j) \right)^{-r_0} \right)^{r_1 / r_0} \right)^{-\frac{1}{r_1}} \tag{11}$$

The parameters are as explained above. With particular preference parameters the *amalgamated* model reduce to the previous functional forms. If $r_1 = r_0$, $b = 0$ and $\tau = 0$, the model is equivalent to the *standard* model. Differing combinations of these simplifications can draw out which of these considerations are important. For notation we use i to denote the individuals within the set N, j within the set n and k for those in set n excluding j. Given the above utility function and the budget constraint $m = \sum_{i=1}^{N} x_i$ the following optimal allocations can be obtained:

$$x_1^* = \frac{m + \sum_{j=2}^{N} \left(\Phi_j \frac{\tau_1}{\pi_j} - \frac{\tau_j}{\pi_j} \right)}{1 + \sum_{j=2}^{N} \left(\Phi_j \frac{\pi_1}{\pi_j} \right)} \tag{12}$$

[4] To relax the assumption of $\tau_i = \tau, \forall i$, giving individual τ_i, if there is only a subset of τ_i where $\tau_{j \neq i} = 0$, then the set of i is reduced to not include j, in the above solution.

$$x^*_{j\neq 1} = \frac{m + \Phi_j^{-1}\frac{\tau_j}{\pi_1} + \sum_{k\neq 1,j}^{N}\left(\frac{\tau_j}{\pi_k}\left(\frac{\alpha'_k\pi_k}{\alpha'_j\pi_j}\right)^{\frac{1}{1+r_0}}\right) - \sum_{k\neq j}^{N}\left(\frac{\tau_k}{\pi_k}\right)}{1 + \Phi_j^{-1}\frac{\pi_j}{\pi_1} + \sum_{k\neq 1,j}^{N}\left(\frac{\pi_j}{\pi_k}\left(\frac{\alpha'_k\pi_k}{\alpha'_j\pi_j}\right)^{\frac{1}{1+r_0}}\right)} \tag{13}$$

Where:

$$\Phi_j = \frac{1}{n^b}\left(\frac{\alpha_0 n^b \pi_j \alpha'_j}{\alpha_1\pi_1}\right)^{\frac{1}{1+r_1}}\left(\sum_{k=2}^{N}\left(\alpha'_k\left(\frac{\alpha'_j\pi_j}{\alpha'_k\pi_k}\right)^{\frac{r_0}{1+r_0}}\right)\right)^{\frac{r_1-r_0}{r_0(1+r_1)}}$$

4 Dirichlet Error Modelling

While the above utility models provide precise *optimal* allocations, x^*_i, for a particular decision problem and preference set, participants are assumed to make 'error' when calculating, or choosing, these allocations. Instead, we assume they draw their *actual* allocations, x_i, from the Dirichlet distribution (Dirichlet, 1839); where the expected values, $E[X_i]$, equal the *optimal* allocations, x^*_i.

The Dirichlet distribution is a multinomial Beta distribution, allowing for N variables, which here correspond to individual allocations (i.e. $x_1, x_2, ..., x_N$), where $x_i \in (0,1)$ and $\sum_{i=1}^{N} x_i = 1$. The below formulates the Dirichlet distribution as a random behavioural model, the work follows from Robson (2017), here altering the variance assumption to allow for varying degrees of complexity, κ.[5] The following assumptions are made: (1) $E[X_i] = x^*_i$, and (2) $Var(X_i) = \frac{(x^*_i(x^*_0 - x^*_i))}{s\kappa^\gamma}$, therefore:

$$E[X_i] = \frac{a_i}{a_0} = x^*_i \tag{14}$$

$$Var(X_i) = \frac{(a_i(a_0 - a_i))}{(a_0^2(a_0+1))} = \frac{(x^*_i(x^*_0 - x^*_i))}{s\kappa^\gamma} \tag{15}$$

Where:

$$a_0 = \sum_{i=1}^{N} a_i, \qquad x^*_0 = \sum_{i=1}^{N} x^*_i$$

It follows that, $\forall i$:

$$x^*_i(s\kappa^\gamma - 1) = a_i \tag{16}$$

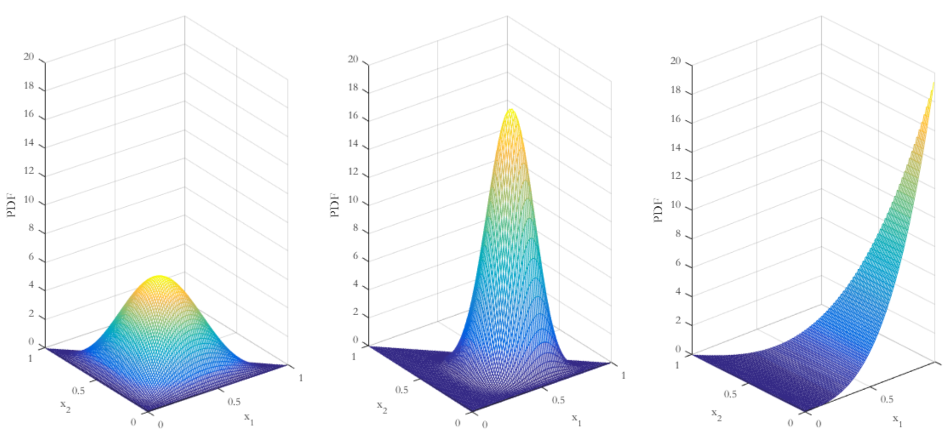
The a_i's determine the shape of the Dirichlet probability density function (*pdf*) and represent the weight given to a particular i. Precision is represented by s, and is multiplied by κ^γ. The

[5]This assumption is relaxed, with two alternative assumptions regarding the variance tested, in Appendix A.5.

higher the value of $s\kappa^{\gamma}$, and therefore the higher α_0, the lower the variance will be. The parameter γ allows for flexibility in the estimation procedure, to identify if *variance* increases or decreases as the degree of complexity, κ, increases, independently of the optimal allocations, where $\gamma \in [-1, 1]$. The degree of complexity, κ, denotes how difficult the decision problem is, by accounting for how many allocation decisions are needed to be made (minus that which is the remainder); here for the 2, 3 and 4 player multiple slider rounds $\kappa = 1, 2, 3$, respectively, while $\kappa = 1$, within the single slider treatment. When $\gamma = 0$, n has no effect on variance, independently of x_i^*, while $\gamma < 0$ implies $s\kappa^{\gamma}$ decreases with κ and if $\gamma > 0$, $s\kappa^{\gamma}$ increases.

To illustrate the above, Figure 2 shows the *pdf*'s of alternative Dirichlet distributions, where $N = 3$, $\gamma = 0$ and $x_3 = 1 - x_1 - x_2$. The left shows an imprecise individual, $s = 10$, who aims to allocate equally $E[X] = [0.33, 0.33, 0.33]$, with $A = [3, 3, 3]$. Second, with $A = [10, 6, 6]$, an individual allocating slightly more to themselves, $E[X] = [0.45, 0.27, 0.27]$, with a greater deal of precision, $s = 23$. Third, with $A = [4, 1, 1]$ more self-interested preferences, here $E[X] = [0.67, 0.17, 0.17]$, can be represented; with a mode where $x_1 \rightarrow 1$ (here precision is low ($s = 7$) but precision can be increased). The flexibility of the Dirichlet distribution is a useful property, and the above derivations allow for easily interpretable parameters to be estimated.

Figure 2: Dirichlet Distribution: Probability Density Function



The preference parameters: α, r, s and γ (alongside r_0, b and τ in their respective models) are estimated, for each individual, through maximising the following log-likelihood function. The preference parameters determine the optimal allocations, x_{it}^*, and consequently the shape parameters, a_{it}, in each round $t \in T$. The multiple integral of the *pdf*, determined by a_{it}, is taken over the n-dimensional 'rounding' interval V_t. V_t is determined by the *observed* decisions, x_{it}; where the 'rounding' interval, around the *observed* decision, is necessary as decisions are not strictly continuous (only to the nearest pence). Estimated parameters are those which maximise the log-likelihood function, hence are the 'most likely' fit for the observed data.

$$\sum_{t=1}^{T} \log \left(\int \underset{V_t}{\cdots} \int \left(\frac{1}{B(a_{0t})} \prod_{i=1}^{N_t} \ddot{x}_{it}^{a_{it}-1} \right) d\ddot{x}_{1t} \ldots d\ddot{x}_{nt} \right) \tag{17}$$

Where:

$$B(a_{0t}) = \frac{\prod_{i=1}^{N_t} \Gamma(a_{it})}{\Gamma\left(\sum_{i=1}^{N_t} a_{it}\right)}, \qquad \ddot{x}_{N_t} = 1 - \sum_{i=1}^{n_t} \ddot{x}_{it},$$

$$V_t = \left\{ (\ddot{x}_{1t}, \ldots, \ddot{x}_{nt}) \in \mathbf{R}^{n_t} \ : \ x_{it} - \frac{0.5}{m_t} \leq \ddot{x}_{it} \leq x_{it} + \frac{0.5}{m_t}, \forall i \in [1, n_t] \right\}$$

The multiple integral is reduced to n dimensions (hence $\ddot{x}_{N_t} = 1 - \sum_{i=1}^{n_t} \ddot{x}_{it}$) as $\sum_{i=1}^{N_t} x_{it} = 1$. This ensures the above condition is met and computational demands are lowered. A penalty function is also applied if $a_i < 0.5, \forall i$, due to the increase in computational demands when calculating triple integrals, at the bounds, when $a_i < 0.5$. In the single slider treatment the number of dimensions of the decision problem is two (hence $\kappa = 1$) for all N, and so allocations to the self (x_1) and total allocations to others (x_o) are modelled, rather than the allocation to each other (x_j). For sample parameter estimates, the log-likelihood contributions for the decisions of every individual within that sample are summed.

5 Results

5.1 Proportional Payoffs

The Proportional Payoff to Player i (PP to Pi), represents the share of payoffs given to a particular player (i). Table 2 shows the mean proportional payoffs to each player, given a particular group size and the type of slider used. In general we observe a decrease in the PP to P1 (the *self*) as the group size increases. On average, participants are willing to sacrifice their own payoffs to increase the total given to the *others*. This increase to others does not, however, maintain the same average level of giving to each *other*. Payoffs between multiple and single sliders are not significantly different (10% level) for any player or group size.

Figure 3 shows the distribution of PP to P1, given the group size and slider type, at the per *round* level (n = 3,611). The top panel shows results from the *multiple* slider, while the bottom panel shows those from the *single* slider. The underlying reasons for the differences in the averages, shown above, emerge. In both panels, as N increases the PP to P1 decreases. The shift in the average is, however, predominantly due to those who are sharing equally between themselves and

Table 2: Average Proportional Payoffs; Players and Sliders

	Multiple Slider				Single Slider	
N Players	P1	P2	P3	P4	PS	PO
2	0.709	0.291			0.712	0.288
3	0.624	0.188	0.188		0.626	0.187
4	0.586	0.139	0.138	0.137	0.604	0.132
6					0.562	0.088
12					0.489	0.046

others. The modal spike at 1 (an average of 23.5% and 26.4% of the sample, for multiple and single sliders, respectively) shifts little as N increases. However, the second model spike, at equal sharing ($\approx 1/N$) shifts proportionately as N increases. Indeed, the same shift is found in both slider treatments.

Figure 3: Distribution of PP to P1: Differing Sliders and N

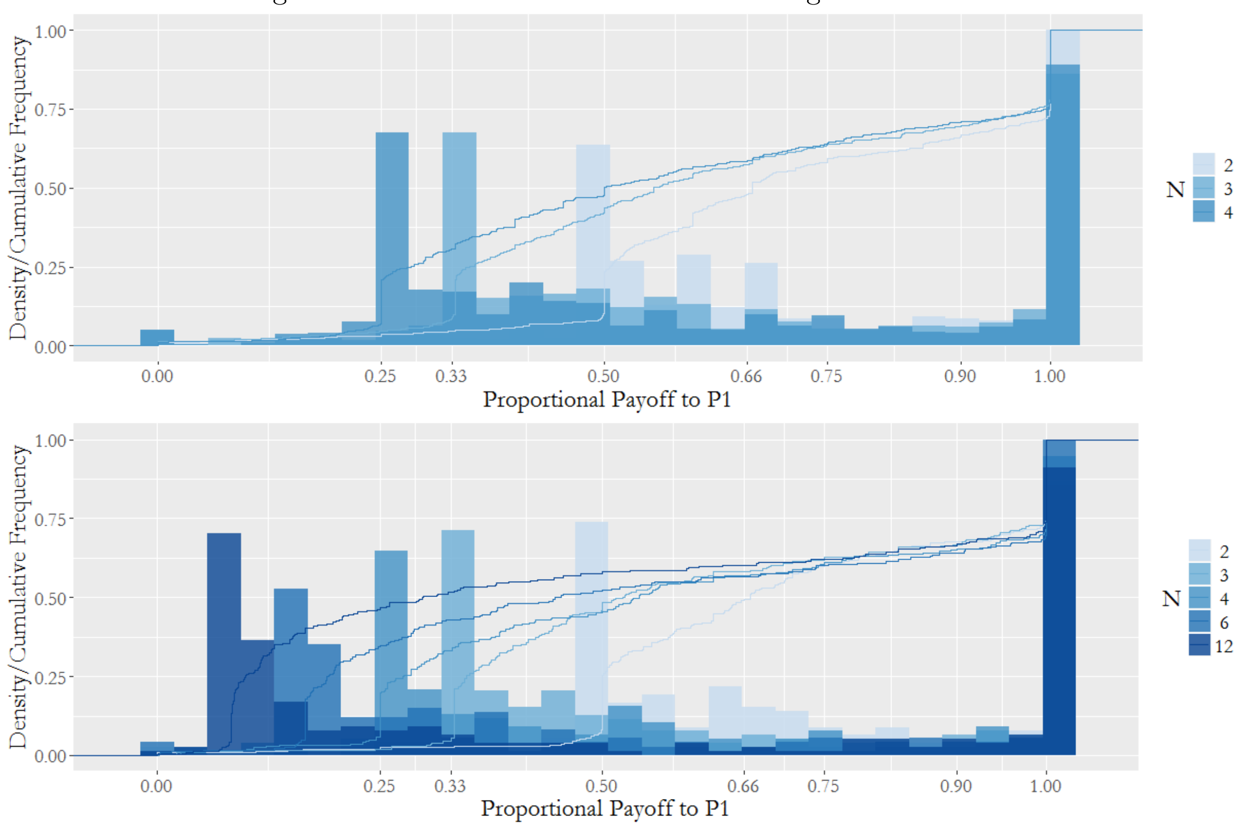


Table 3 shows results from two random effects models, with the Proportional Payoff to Player 1 (the *self*) and Player j (each *other*) as dependent variables. Results show that as N increases there are large and significant effects on the PP to P1 and Pj. With $N = 2$ as the reference category, we observe that an increase in the number of others leads to a reduction of the PP to P1 and a decrease in the PP to Pj. This reveals that, on average, participants are willing to significantly reduce their own payoffs, and therefore increase the total payoffs to others, but not to the extent that the average payoffs to others remain constant. We observe that switching from the multiple to single slider has no significant effect on behaviour, and neither does the round number. The relative multiplier for P1 ($\pi_1/(\sum_{j=2}^{N} \pi_j/n)$) is included as a control, and is positively correlated. An increase in the budget, standardised within each N, is shown to have a negative effect on PP to P1, and a positive effect on PP to Pj. Neither the sign nor significance of any coefficient changes when those who on average keep more than 0.99 of the proportional payoffs to themselves are excluded from the analysis. The same is true when an extensive list of demographic characteristics (excluding parental income), 'oneness' levels and opinion questions are included as controls; with the exception of the significance levels of the budget levels, which decrease.

Table 3: Random Effects Model: Proportional Payoff to P1

	(1) **PP to P1**		(2) **PP to Pj**	
	Coef.	Std. err.	Coef.	Std. err.
N Players				
- 3	-0.0717***	(0.0103)	-0.1100***	(0.0092)
- 4	-0.1037***	(0.0134)	-0.1625***	(0.0116)
- 6	-0.1493***	(0.0203)	-0.2023***	(0.0146)
- 12	-0.2137***	(0.0245)	-0.2476***	(0.0172)
Single Slider Dummy	0.0092	(0.0163)	0.0050	(0.0095)
Relative Multiplier P1	0.0560***	(0.0077)	-0.0316***	(0.0046)
Standardised N Budget	-0.0270**	(0.0119)	0.0124**	(0.0057)
Round Number	-0.0003	(0.0006)	-0.0002	(0.0004)
Constant	0.6406***	(0.0244)	0.3354***	(0.0213)
N	83		83	
Observations	3611		5985	
R^2 Within	0.2019		0.2659	
R^2 Between	0.0023		0.1044	
R^2 Overall	0.0839		0.1667	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Further analysis of the design parameters, including the player name (i.e. 2, 3, 4), screen position (i.e. left, middle, right) and the randomised order of N (all of which are found to have insignificant effects) are found in Appendix A.2, alongside analysis of the effects of demographic characteristics, 'oneness' levels and opinion questions on giving behaviour.

6 Preference Parameters

While the above analysis describes the observed aggregate-level behaviour and treatment effects, it lacks an explanation of why such behaviour is observed. Behaviour depends on preferences, and aggregate-level behaviour ultimately depends on the nature and distribution of individual preferences. Through estimating preference parameters, by assuming participants are behaving *as if* they are (*noisily*) maximising a utility function, the preferences held by participants can be characterised. This characterisation enables intuitive insights into the reasons why we observe such behaviour. Below, aggregate-level preference parameters are estimated to characterise the preferences of the representative agent. Then, individual-level preference parameters are estimated; the distribution of which accounts for the aggregate trends observed.

6.1 Aggregate Preference Parameters

At the aggregated level preference parameters can be estimated for a representative agent, within each of the utility functions proposed. The following results characterise how the sample behaves on average, but also identify how additional preference parameters affect the estimates of those in the simpler models. Table 4 shows the estimated preference parameters and parameters within the error model.

Results from the *standard* model shows $\alpha = 0.328$ and $r = 0.143$, showing a high regard for others and weakly weighted prioritarianism. Self-interest parameters are similar for both *FKM* and *andreoni* models, but are lower than those estimated in the *stone-geary* and *amalgamated* models. This difference is perhaps explained by the inclusion of τ, as behaviourally individuals will allocate equally until each player have more payoffs than the minimum threshold, and then distribute according to their α (and other parameters).

We observe that the estimates for r are lower than the estimates of r_1 value of 6.652, within the *FKM* model. This difference perhaps accounts for the split of r into r_1 and r_0, showing that individuals are more averse to inequality between themselves and others, but will slightly prioritise efficiency between individuals. Estimates of b, show that participants will increase the total payoffs to others as N increases, but will not maintain the same average payoff. The inconsistency within the estimated parameters lie in the estimated r_1 in the *FKM* and *amalgamated* models, where the r_1 in the latter implies efficiency prioritisation between the self and others.

Estimates of the error parameters, are relatively consistent. We observe low values of s, which are expected due to the pooled nature of the data. The elasticity of precision, γ, is positive and relatively high in each estimation. This implies that as the degree of *complexity* increases the variance of X_i will decrease.

Table 4: Sample Level Estimates of Parameter Values

	Preference Parameters					**Error Parameters**	
	α	r	r_0	b	τ	s	γ
Standard	0.328	0.143	.	.	.	3.547	0.600
FKM	0.331	6.652	-0.093	.	.	3.880	0.516
Andreoni	0.346	0.685	.	0.821	.	4.802	0.547
Stone-Geary	0.492	0.793	.	.	2.451	2.927	0.958
Amalgamated	0.490	-0.594	-0.087	0.515	5.839	4.319	0.581

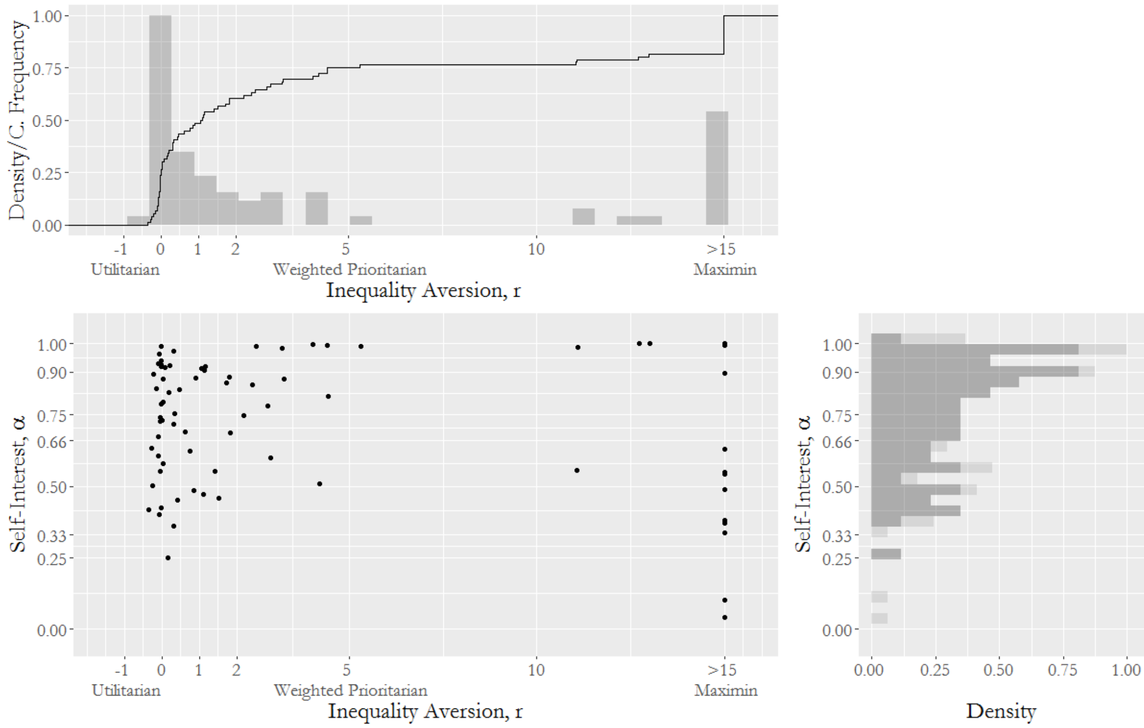
6.2 Individual Level

Preference parameters are estimated at the individual level, for each of the five models. For the analysis that follows there are seven individuals who are excluded, as they made purely self-interested decisions in every round. They are classed as 'egoists' who have $\alpha = 1$. The remaining 76 participants have individual-level preference parameters estimated.

Figure 4 shows the distribution of inequality aversion, r, and self-interest, α, estimated with the *standard* model. The top-left and bottom-right panels show histograms (and cdf plot) of r and α, respectively, while the bottom-left shows a scatter plot of the two variables. According to r the individuals classified into five different categories. There are 22.37% classified as 'Efficiency Prioritarians' ($r < -0.01$), 4.95% exhibit preferences close to 'Cobb-Douglas' ($-0.01 \geq r \leq 0.01$), 55.26% who are 'Weighted Prioritarians' ($0.01 > r < 15$) and 18.42% who are 'Maximin' ($r \geq 15$). The median value of $r = 1.08$. As r increases α becomes increasingly difficult to interpret, as a result the histogram of α shows a stacked histogram, where the lighter grey plot shows the distribution of α where $r > 10$ and the darker plot where $r <= 10$. Of those, where $r <= 10$, 15.5% have $\alpha < 0.5$, 44.8% have $\alpha < 0.75$ and 70.7% have $\alpha < 0.9$. Note that, in addition to this, seven 'egoists' are omitted, who have $\alpha = 1$. These results show that, the majority of the sample have a substantial regard for others and are willing to sacrifice total payoffs in order to increase the payoffs of the worst-off. Yet there are significant minorities within the sample who are predominantly self-interested alongside others who prioritise efficiency.

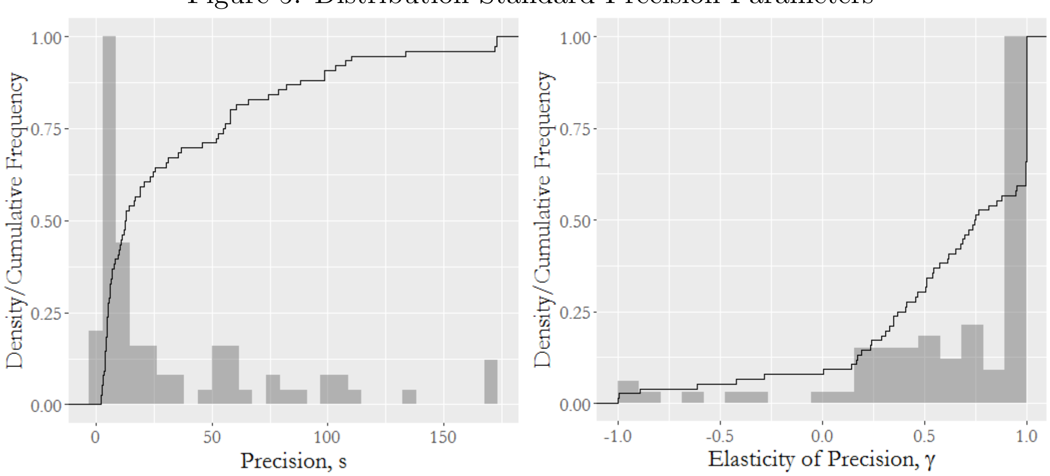
Alongside preference parameters, the error parameters s and γ are estimated for each individual. Figure 5 shows the distribution of the two parameters. The left panel shows s, the precision parameter. The higher is s the lower the variance X_i. There are 23.68% of the sample with $s \leq 5$, 42.11% with $s \leq 10$ and 71.05% with $s \leq 50$. Elasticity of precision, γ, identifies how the variance of X_i changes as N increases; if $\gamma < 0$, ceteris paribus, variance decreases as N increases, while if $\gamma = 0$ there is no change, and if $\gamma > 0$ there is an increase. The right panel show the distribution of γ. Only 7.89% of the sample have $\gamma < 0$, $\gamma > 0$ for 92.11%, and $\gamma > 0.99$ for 40.79%. These results imply that, for the majority, as N increases, the variance of X_i decreases. In other words,

Figure 4: Distribution Standard Preference Parameters



individuals draw their actual allocations, x_i, closer to the optimal allocations, x_i^*, more frequently as the number of recipients increases.

Figure 5: Distribution Standard Precision Parameters



6.2.1 Between Model Parameter Comparisons

Within each of the five utility functions preference parameters for inequality aversion, r, and self-interest, α, are estimated, alongside other parameters of interest. Of interest, is the difference between the estimated parameters, as the incorporation of alternative (potentially omitted) parameters may effect the estimates. Table 5 shows the p-values of a one-sided Sign-Test of Matched Pairs, between pairs of estimates from alternate utility functions. The test is used as it accounts for the matched nature of the data, and makes no assumptions about the distribution of the parameters. The null hypothesis is that the median of differences, between the parameters, is zero. The alternative hypothesis is that the median of the difference in parameters is less than zero. A low p-value, therefore, rejects the null, showing that the parameters estimated from the first model (denoted in the column) are lower than the second model (denoted in the row).

Table 5: Between Model Comparison of r and α: Sign-Test of Matched Pairs (p-values)

	Self-Interest, α				
	Std	Fis	And	SG	Amal
Std	1.000	0.849	0.789	0.849	0.634
FKM	0.211	1.000	0.634	0.546	0.546
And	0.283	0.454	1.000	0.546	0.546
SG	0.211	0.546	0.546	1.000	0.454
Amal	0.454	0.546	0.546	0.634	1.000

	Inequality Aversion, r				
	Std	Fis	And	SG	Amal
Std	1.000	0.634	0.717	1.000	0.986
FKM	0.454	1.000	0.634	1.000	0.998
And	0.366	0.454	1.000	1.000	1.000
SG	0.000	0.000	0.000	1.000	0.151
Amal	0.025	0.004	0.001	0.897	1.000

	Precision, s				
	Std	Fis	And	SG	Amal
Std	1.000	0.008	0.151	0.849	0.004
FKM	0.996	1.000	0.932	0.986	0.008
And	0.897	0.103	1.000	0.789	0.000
SG	0.211	0.025	0.283	1.000	0.000
Amal	0.998	0.996	1.000	1.000	1.000

	Elasticity of Precision, γ				
	Std	Fis	And	SG	Amal
Std	1.000	0.634	0.789	0.454	0.897
FKM	0.454	1.000	0.789	0.283	0.932
And	0.283	0.283	1.000	0.366	0.634
SG	0.634	0.789	0.717	1.000	0.932
Amal	0.151	0.103	0.454	0.103	1.000

Results show that there are no significant differences between estimates of either α or γ, for any pair of models. There are also no significant differences in r between the *standard*, *FKM* and *andreoni* models. We do, however, observe that the estimates of r for *stone-geary* are significantly lower than those in the *standard*, *FKM* and *andreoni* models. Likewise the estimates for the *amalgamated* model are shown to be significantly lower than those in the *standard*, *FKM* and *andreoni* models. This difference is apparent when considering the proportion of the sample classed as 'maximin' ($r \geq 15$), by each model; with only 9.3% and 7.9% for *stone-geary* and *amalgamated*, respectively, compared to 18.4%, 14.5% and 17.1%, for *standard*, *FKM* and *andreoni* respectively. The *stone-geary* and *amalgamated* estimates of r are not significantly different. These results imply that the inclusion of τ partially accounts for the equal-sharing behaviour, which the other models explain through a higher r.

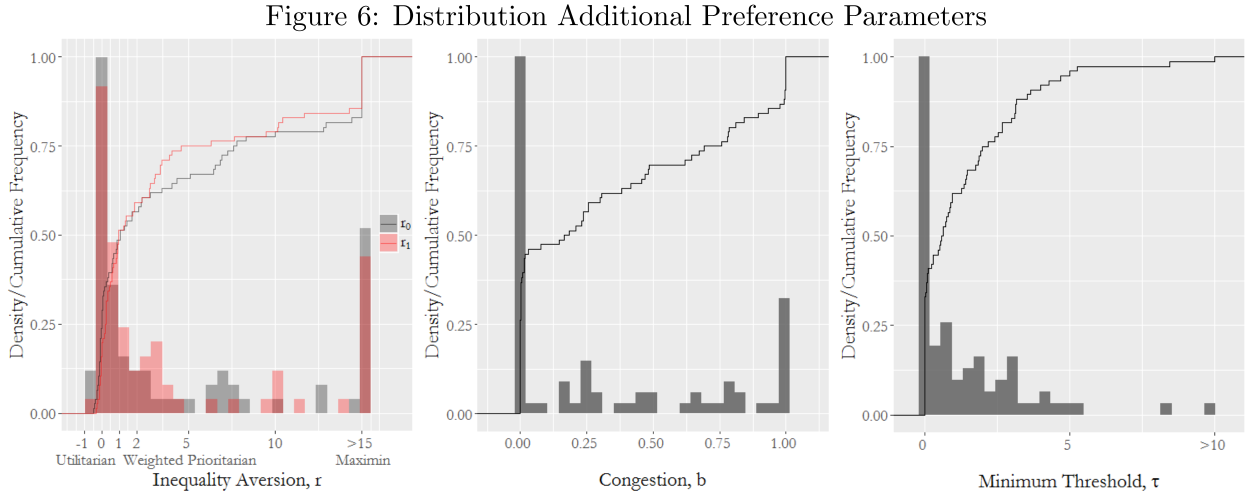
Results from comparisons of s show that the parameters estimated from the *standard* model are higher than those in the *FKM* model, and those estimated in the *stone-geary* model are lower than those in the *FKM* model. The parameters within the *amalgamated* model are significantly higher than those in each of the other models, indicating that the additional parameters within the model allow for the flexibility for a more precise fit.

6.3 Extended

In addition to the preference parameters described above, the extended preference parameters r_0, r_1, b and τ are estimated; the distributions of which are shown in Figure 5. The left panel shows the distribution of r_0 and r_1, estimated from the *FKM* model, the middle panel shows b, from the *andreoni* model, and the right, τ, from the *stone-geary* model.

The distributions of r_0 and r_1 show potential differences in *self-other* and *between-other* equality-efficiency trade-offs. The distributions are similar, but r_0 tends to take more extreme values. 23.68% of the sample have $r_0 < -0.01$, compared to 15.79% for whom $r_1 < -0.01$. Similarly, $r_0 \geq 15$ for 17.11%, while 14.47% have $r_1 \geq 15$. The two preferences are strongly correlated, with a spearman's rank correlation coefficient of 0.786. Those who have r_0 and r_1 with the same sign make up the majority of the sample; for 53 individuals $r_0, r_1 \geq 0$, while $r_0, r_1 < 0$ for 9 individuals. There are, however, 4 for whom $r_0 \geq 0$ and $r_1 < 0$ and 10 for whom $r_0 < 0$ and $r_1 \geq 0$.

Figure 6: Distribution Additional Preference Parameters



The distribution of b shows that the 'average' and 'total' payoffs to others matter for different individuals. For a large proportion of the sample (39.47%) it is 'average' payoffs which matter ($b < 0.01$), however, a significant amount (13.16%) consider the 'total' payoffs ($b > 0.99$) and do not reduce the payoffs to the self as n increases. Those who have a parameter between 0.01 and

0.99 make up the remaining 47.37%. The mean value of b is 0.336. Minimum thresholds, τ, also vary between individuals. For 73.68% $\tau > 0.01$, $\tau > 1$ for 38.16%, $\tau > 3$ for 17.11% and $\tau > 5$ for 5.26%. Showing that for the majority of the sample there is a minimum threshold which they will allocate before considering other self-other and equality-efficiency trade-offs. The median value of τ is $62p$.

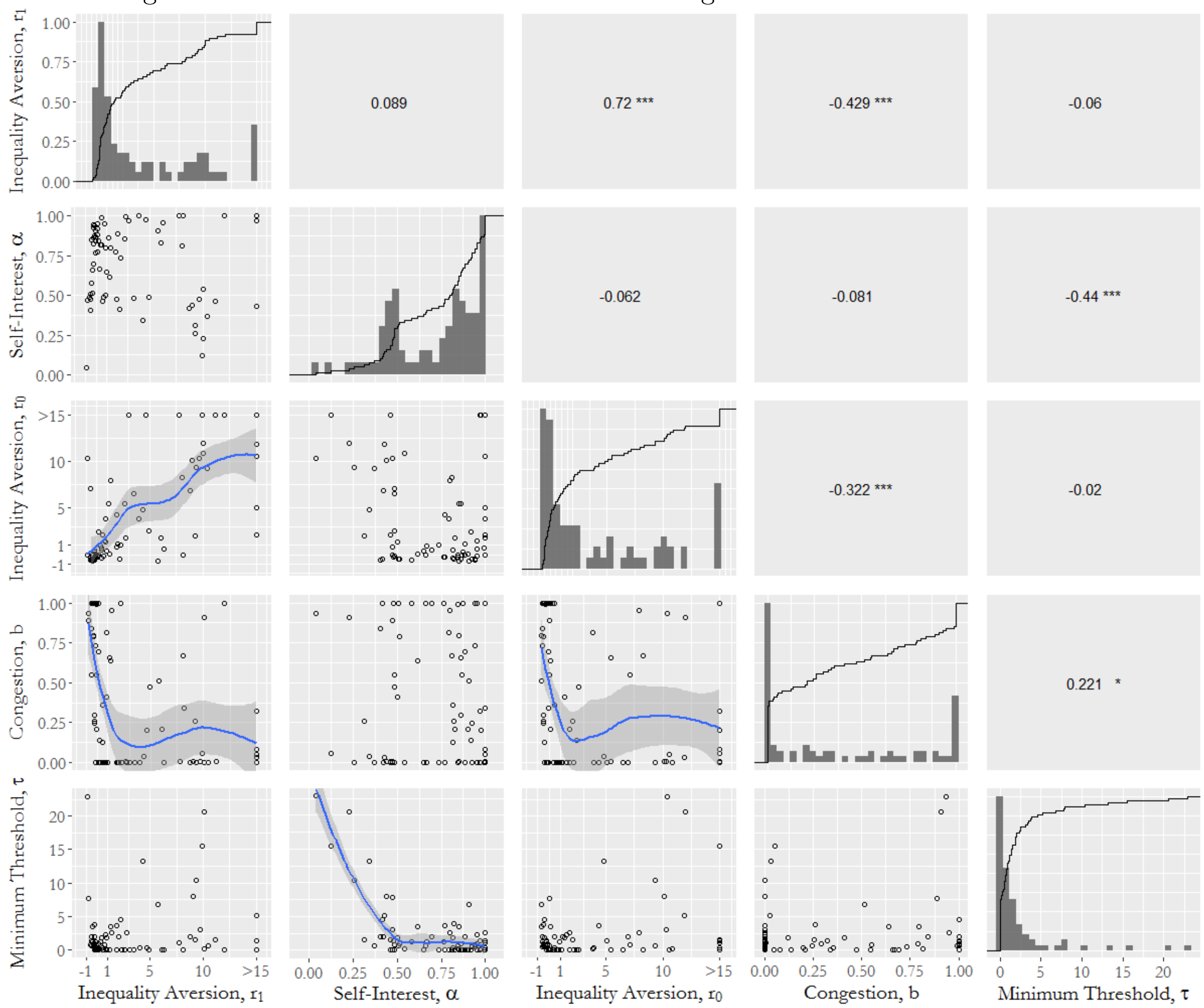
The differences in r_1 and r_0, alongside parameter values of b and τ show that the behaviour of participants diverges from that predicted in the *standard* model. Indeed, by performing likelihood-ratio tests (to identify if the additional preference parameters increase the goodness-of-fit significantly, at the 10% level) individuals can be separated into either *standard* or (one of the) *extended* types. The preference parameters (estimated from the *extended* model) can then be compared. We observe a median τ of 1.881 for those the *stone-geary* model fits best, with a lower 0.127 for those for whom the *standard* model fits best. The median congestion, b, is 0.672 and 0.000 for those in the *andreoni* and *standard* models respectively. For comparing r_1 and r_0, we are interested in the difference between the parameters, so calculate the weighted euclidean distance from the estimated (r_0, r_1) to the closest point on the line where $r_0 = r_1$ as: $d = (((r_0 - r_1)/2)^2 + ((r_1 - r_0)/2)^2)^{1/2}/((|r_1| + |r_0|)/2 + 1)$. We observe a median distance, d, of 0.616 and 0.220, for those in the *FKM* and *standard* models, respectively. The preference parameters which ensure the *extended* model diverges most from the *standard*, are observed to a greater degree when the *extended* models fit individual behaviour better than the *standard* model.

6.4 Amalgamated

Of interest is not only the distribution of individual level preference parameters, but their relation with one another. Figure 11 shows the histograms (combined with cumulative frequency plots) of each preference parameter on the diagonal. In the bottom-left triangle, are the scatterplots of each corresponding pair of preference parameters. The top-right triangle shows Spearman's rank correlation coefficients. Loess (local regression) fitted curves are shown (with 95% ci) if coefficients (from the mirrored panel) are significant at the 5% level. The distribution of preference parameters, is somewhat similar to those from the above individual utility functions. As shown in the between model comparisons, there are no significant differences in α, but $r(_1)$ is significantly lower.

Correlations between preference parameters can also be established. First, inequality aversion parameters, r_0 and r_1, are positively correlated (0.72^{***}); participants between-other preferences appear to be closely related to their self-other preferences. Congestion, b is negatively correlated with both r_1 and r_0 (-0.429^{***} and -0.322^{***}, respectively), this implies that as the number of others increases, those who are more efficiency seeking would sacrifice their own payoffs to maintain the total to others. The minimum threshold, τ, is negatively correlated to α, -0.44^{**}; this result

Figure 7: Distribution and Correlation of Amalgamated Preference Parameters



is partially driven by the few individuals with very high τ, as they are almost precisely equally distributing payoffs equally each round.

To summarise the estimated preference parameters, and to relate the estimations to the increased complexity that the amalgamated model incorporates, Table 6 tabulates individuals for whom the *extended* parameters are 'negligible' or not. The 'negligible' *extended* preference parameters are those which would collapse the amalgamated function to a more simple functional form. Those classed as 'negligible' are when $b \leq 0.01$, $\tau \leq 0.1$ and $d \leq 0.25$, the eight possible combinations of parameters being 'negligible' or not are shown in Table 6. The top-left results shows that all parameters are 'negligible' for 3 individuals, while the bottom-right shows 14 individuals for whom all parameters are 'non-negligible'. Results between the two extremes show the combinations of which *extended* preferences are important. There are 36 for whom differences in r_0 and r_1 are large enough, 47 for whom $b > 0.01$ and 52 for whom $\tau > 0.1$. For those with only $d > 0.25$ and $b > 0.01$ there are 8, only $d > 0.25$ and $\tau > 0.1$ there are 6, while there are 20 with only $b > 0.01$ and $\tau > 0.1$. This heterogeneity points to models which could distinguish between either having any one, a combination of two or all three *extended* preference parameters accounted for.

Table 6: Summary of Amalgamated Preference Parameters

		Minimum Threshold, τ				
		$\tau \leq 0.1$ **Congestion, b**		$\tau > 0.1$ **Congestion, b**		Total
		$b \leq 0.01$	$b > 0.01$	$b \leq 0.01$	$b > 0.01$	
Inequality Aversion, d	$d \leq 0.25$	3	5	12	20	40
	$d > 0.25$	8	8	6	14	36
Total		11	13	18	34	76

Further analysis in Appendix A.6 uses a finite mixture model to identify 'clusters' of individuals. This allows for an intuitive summary of the high dimensional preference parameters, characterising groups of participants with similar preferences.

7 Goodness-of-Fit and Predictive Accuracy

Analysis can be conducted on both goodness-of-fit and predictive accuracy to determine how well the utility functions proposed explain individual behaviour. The 'best' utility model can be identified for each individual, splitting the sample into different 'types'. The alternative utility functions can be ranked, by comparing the maximised log-likelihood (MLL) values. The MLL is a measure which accounts for the stochastic nature of individual behaviour, as the measures are constructed of the likelihood of observing the actual behaviour, given the preferences estimated and error model assumed.

Due to the experimental design both *goodness-of-fit* and *predictive accuracy* can be analysed. MLL values can be calculated for *multiple slider* treatment on which the preference parameters are estimated, determining goodness-of-fit, and for the *single slider* treatment, using those estimated parameter values, to determine predictive accuracy. The ability of a model to both fit and predict behaviour is important, therefore, analysis of the two separately and as a combined measure 'Both' (a weighted average of the two) is conducted to identify if particular models are 'best' in either criteria.

An issue with comparing the 'raw' MLL is that alternative models may have a differing number of parameters. Models with a larger number of parameters are more flexible so should fit behaviour better; yet, if the difference is small the additional complexity of the model is perhaps not warranted. Several measures of *information criterion* seek to address this trade-off between fit and model complexity. Three commonly used alternatives are the Akaike information criterion (AIC) (Akaike, 1998), Bayesian information criterion (BIC) (Schwarz et al., 1978) and Hannan–Quinn information criterion (HQI) (Hannan and Quinn, 1979).[6] The three criterion may give slightly alternative rankings, due to differences in their calculation and different implicit trade-offs being made between the fit and model complexity. To sidestep such differences, the three criterion are calculated for each of the five models, for each individual, and a composite criterion, the *information criterion* (IC), is constructed whereby a model is 'best' if two or more of the criteria rank that model highest.

Table 7 tabulates the above. Results from the MLL are shown to the left of the IC, each split into three columns: goodness-of-fit, predictive accuracy and both. The results show the importance of comparing goodness-of-fit and predictive accuracy, as well as accounting for the trade-off between fit and complexity, as mismatches in the rankings occur. The *amalgamated* model shows this most starkly. In the MLL GOF it is the modal 'type', with 29 individuals for whom it fits 'best'. This number drops to only six and seven in predict and both, respectively. The higher number of parameters allows the flexibility to fit data well, but this comes at a cost of predictive power. Furthermore, when penalising the function for the higher number of parameters the information criteria finds there are no individuals for whom the *amalgamated* function is 'best', in either 'Predict' or 'Both'. Results are opposite for the *standard* model, there are less for whom the model is 'best' in GOF compared to 'Predict', and in MLL compared to IC. The three models with five parameters, tend to lie somewhere in between these extremes.

The results of most interest are in the final column. These rankings are those which will be used to determine the 'type' of each individual. The modal 'type' is the *standard* model and no individuals are classed within the *amalgamated* model. A substantial proportion of the sample are classed as *extended* types, with 13 *FKM* types, 21 *andreoni* type and 9 *stone-geary* types.[7]

7.1 Likelihood Proportions and R^2

While the above analysis shows how the models do relative to one another, it reveals little about how well the model performs in absolute terms. The standard metric to analyse performance of

[6] The information criteria statistics are as follows: AIC $= 2k - 2(MLL)$, BIC $= \ln(n)k - 2(MLL)$ and HQI $= 2k.\ln(\ln(n)) - 2(MLL)$, where k = number of estimated parameters and n = number of observations.

[7] Appendix A.7 discusses and analyses mismatches between rankings; firstly by using RSS and secondly with preferences estimated using alternative error modelling.

Table 7: Utility Types: Ranked by Log-Likelihood and Information Criterion

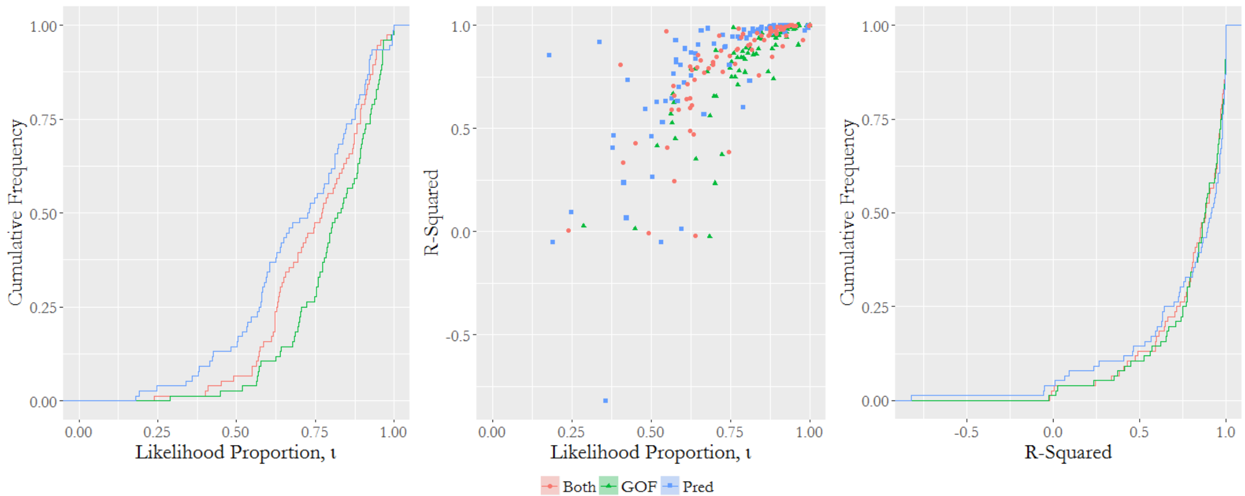
	Log-Likelihood			**Information Criterion**		
	GOF	Predict	Both	GOF	Predict	Both
Standard	9	16	19	27	41	33
FKM	14	17	17	11	11	13
Andreoni	21	24	23	26	19	21
Stone-Geary	3	13	10	9	5	9
Amalgamated	29	6	7	3	0	0

a model is R^2, which determines how much of the sample variation in the variable of interest, is explained by the model. The *likelihood proportion*, ι, is an alternative metric, which focuses on likelihood contributions. In each decision problem, t, the *likelihood contribution*, l_t, is calculated as the area under the probability density function (given by estimated preference and error parameters) within the 'rounding' interval, around the *observed* decision (see Section 4). Intuitively, l_t, denotes the likelihood of observing the decision made, given the error model. The uniform likelihood contribution, lU_t, can likewise be derived from assuming that the probability density function takes the form of a uniform distribution. This denotes the likelihood of observing the decision made, given uniformly random draws are made. The *likelihood proportion* in each decision problem is defined as, $\iota_t = l_t/(l_t + lU_t)$. The measure shows how much 'more likely' the observed behaviour is in the specified model, in relation to the uniform distribution. If $\iota_t > 0.5$ the proposed model does 'better' at explaining behaviour than uniformly random draws, if $\iota_t = 0.5$ then the two are equal, while if $\iota_t < 0.5$ the uniform distribution 'better' explains behaviour. The summary measure $\iota = \sum_1^T(\iota_t)/T$ shows how well the proposed model explains behaviour, on average.

Figure 8 shows the distribution of ι, in the left panel, and R^2, in the right, across individuals in the sample. The model assumed for each individual is that based on their 'type', established in Table 7. The measures are calculated for each the *goodness-of-fit*, *predictive accuracy* and *both*. For both ι and R^2 the measures within the goodness-of-fit measures tend to be higher than the predicted accuracy, with both lying between. The mean values for $\iota = 0.806, 0.696$ and 0.751, for GOF, pred and both, respectively, with mean $R^2 = 0.806, 0.770$ and 0.796, respectively. There are 2, 11 and 5 individuals for whom $\iota < 0.5$ and 1, 3 and 2 for whom $R^2 < 0$, for GOF, Pred and Both, respectively. The central panel shows a scatter plot of ι and R^2, highlighting the strong correlation between the two measures (with Spearman's rank correlation coefficients of 0.90, 0.88 and 0.87 for GOF, Pred and Both, respectively).

The two measures similarly aim to measure the strength of the models proposed. R^2 focuses on how close observed decisions are to the optimal decisions proposed by the utility function. The *likelihood proportion*, ι, however, incorporates the stochastic assumptions made, identifying how more often the model proposed would predict the observed behaviour. Both measures show that

Figure 8: Distribution of ‘Best’ Likelihood Proportions and R^2



for the majority of the sample, the models proposed and preferences estimated explain well the observed behaviour.

8 Discussion

8.1 Comparing Giving

Results from the *single slider* treatment are particularly comparable with those from Andreoni (2007). Table 8 shows the mean PP to P1 and PP to PO, where PO represent the average payoff to others, for differing N.[8] Results show that in the two-player game participants in our experiment are less generous, than those in Andreoni (2007). As N increases, however, while participants give a lower proportion to themselves in our results, the PP to P1 do not decrease, and indeed appear to have an upward trend, in Andreoni. The PP to PO follow a similar, but opposite, trend, with the average PP to others being approximately equal in our twelve-person treatment as the six-person treatment in Andreoni (2007).

While these differences are interesting, they should be approached with caution. The distributional decisions of participants is heavily dependant upon the experimental design parameters, the particular choice of dividers/multipliers, budgets and incentives will have differential effects on raw giving, depending on the preferences of participants. Indeed, this is one important reason why estimating preference parameters is important; if preferences are estimated then behaviour

[8] Results from Andreoni (2007) are calculated from individual level data from: http://econweb.ucsd.edu/ jandreon/WorkingPapers/GARPN%20cesEstimates%20APX%20table.htm

in differing experimental designs can be predicted to identify differences not purely based on experimental design. One difference between the designs is the difference in average budgets as N increases; within our design the average budget remains the same, while in Andreoni's design it decreases. Similarly, the incentive structure leads to different behaviour. In Andreoni's set-up the participant knows they will receive the payoff they give to themselves, plus the 'Pass' payoffs from each of the n other participants in their group.

Table 8: Comparing Average Proportional Payoffs

	Andreoni (2007)		**Robson**	
N Players	P1	PO	P1	PO
2	0.622	0.378	0.712	0.288
3	0.710	0.145	0.626	0.187
4	0.688	0.104	0.604	0.132
5	0.695	0.076	.	.
6	0.756	0.049	0.562	0.088
10	0.727	0.030	.	.
12	.	.	0.489	0.046

Results from Fisman, Kariv, and Markovits (2007), do however, appear to be more in line with our results. Comparing results to the *multiple slider* treatment, the equivalent mean PP to P1 is 0.79 and 0.75, in the two and three person treatments, of their experiment. Comparing this to our 0.71 and 0.62, we observe that 'dictators' take less for themselves as N increases; however, both the absolute level of generosity and the change in giving are higher in our experiment.

8.2 Comparing Preference Parameters

8.2.1 Self-Other and Between-Other Inequality Aversion

While Fisman, Kariv, and Markovits (2007) (FKM) run both two and three-person dictator games, preference parameters are estimated separately for each treatment. We compare classifications of r_1 and r_0 with those estimated in their three-person treatment, and the r estimated in their two-person treatment. To make estimates comparable, we use their classifications, and exclude those 'selfish' individuals with an average PP to P1 greater than 0.95 or who are not 'consistent'.[9] Our total sample of participants with "consistent nonselfish preferences" is 63, with 33 from the three-person and 45 from the two-person treatments in FKM; the percentages shown below in Table 9 refer to these totals.

[9] In their paper they calculate Afriat's Critical Cost Efficiency Index (CCEI) and exclude those individuals with $CCEI < 0.8$, as they behave in a manner 'inconsistent' with utility maximisation. We do not calculate CCEI values, but instead use the likelihood-proportion value, ι to exclude those with $\iota < 0.5$; which (while it is a test dependent upon the utility function chosen) excludes individuals for whom random behaviour better explains their behaviour.

Results in Table 9 show the categorisation of inequality aversion parameters in FKM and this study. We observe that for FKM the majority of the sample are either 'utilitarian' or 'efficiency prioritarians' for both r_1 (66.7%) and r_0 (66.7%), although there is a lower proportion within this categorisation in the two-player experiment for r (53.3%). The opposite is true from our results, with the majority of the sample being either 'weighted prioritarians' or 'maximin', for r_1 (85.7%) and r_0 (74.6%). This reversal shows a much higher weight on *efficiency* concerns for the FKM sample, in contrast to a higher concern for *equality* in our sample.

Table 9: Comparison of Inequality Aversion, r_1 and r_0

	FKM			**RB**	
	r (2P)	r_1 (3P)	r_0 (3P)	r_1	r_0
Utilitarian ($r < -0.9$)	4.4%	12.1%	15.2%	0.0%	0.0%
Efficiency Prioritarian ($-0.9 < r < -0.1$)	48.9%	54.6%	51.5%	6.4%	9.5%
Cobb-Douglas ($-0.1 < r < 0.1$)	11.1%	3.0%	9.2%	7.9%	15.9%
Weighted Prioritarian ($0.1 < r < 0.9$)	31.1%	21.2%	6.1%	22.2%	14.3%
Maximin ($r > 0.9$)	4.4%	9.0%	18.2%	63.5%	60.3%

There could be several reasons for these differences. The first, is the sample. Participants in the UK are perhaps more averse to inequality than their US counterparts. The second, the differences in experimental design alter individual behaviour. In our design participants had to individually allocate to each individual, with a slider, while in their design a single point on a budget line was clicked. The latter allows for quicker and easier decisions to be made, while the former requires more effort. In itself, this could lead to different responses; on the one hand the former method could lead to more 'considered' distributions, accounting for each of the other participants, on the other the ease of clicking a single point could allow for more time to consider the efficiency implications of the choices made. This, however, should then appear in the distributional decisions between the *single* and *multiple* slider treatment, which it does not.

A further difference in design, is the incentive structure. In our design one 'dictators' choice is picked at random to determine the payoffs of all in the group, while in FKM each participant receives the payoffs they gave to themselves, plus the payoffs others gave them. This may have an impact on average giving, but also on trade-offs between equality and efficiency. On average, participants know that if everyone distributes efficiently then payoffs will be greater, but in FKM this carries a much lower risk of particular individuals receiving a low payoff. Other difference include: the explicit statement of the 'Payoff Gap' and 'Total Payoffs' (representing the trade-off between equality and efficiency) in our design; the explicit statement of the 'Dividers' opposed to the difference in graphical representation; and the difference in language between 'allocations' to each player (implying the budget is a common good) compared to 'hold' and 'pass' (implying the budget belongs to the 'dictator', which they can choose to share).

Consistent between our findings is that there are strong within-subject correlation between r_1 and r_0. With FKM there was 63.6% of the sample with $r_0, r_1 \leq 0$, while $r_0, r_1 < 0$ for 24.2%. With only 6.1% with $r_0 \geq 0$ and $r_1 < 0$ and 6.1% with $r_0 < 0$ and $r_1 \geq 0$. In our (similarly reduced) sample there are 81.0% of the sample with $r_0, r_1 \geq 0$, while $r_0, r_1 < 0$ for 4.8%. With 4.8% with $r_0 \geq 0$ and $r_1 < 0$ and 9.5% with $r_0 < 0$ and $r_1 \geq 0$. This means that there are 87.9% and 85.7% of the sample, for FKM and our study respectively, with both *self-other* and *between-other* inequality aversion in the same direction.

8.2.2 Congestion

Andreoni (2007) estimates the congestion parameter, b, at both the sample and individual level. At the sample level, the representative b estimated was 0.68, which is slightly lower than our estimated value of 0.82, but not extensively so. At the individual level Andreoni (2007) estimates preferences for 109 participants, with 11 participants identified as 'perfectly selfish'. Of those 109 participants $b = 0$ was estimated for 25%, while $b = 1$ for 17% and $0 < b < 1$ for the remaining 58%. From our estimates, there are 39.5% of the sample for whom $b < 0.01$, 13.2% with $b > 0.99$ and the remaining 47.4% with $0.01 \leq b \leq 0.99$. The results are somewhat similar, spikes at either extreme, where the modal group has $b \to 0$; but the majority exhibits some degree of congestion.

8.2.3 Minimum Threshold Levels

Comparison with the τ preference parameter within the Stone-Geary function is limited. Its use is more common in other literatures, such as the time and risk preferences. Andreoni and Sprenger (2012) estimate Stone-Geary "consumption minima" (ω_1), within a CRRA utility function with quasi-hyperbolic discounting. While contextually different, the experimental set-up is somewhat similar, with convex time budgets. Their aggregate estimate of $\omega_1 = \$1.35$, when $\omega_1 = \omega_2$ is assumed (the hypothesis of which is not rejected), which lies somewhere between our median individual estimate of 62p and aggregate estimate of £2.45. Of interest, however, is that they find the estimates of other preference parameters (especially curvature) depend on the assumed ω_i, a result we also find (with significantly lower estimates of r in the *stone-geary* in comparison to the *standard* model, in Table 5). Andersen et al. (2008) also use a similar functional form, but do not estimate a minimum threshold, instead utilising the average value of daily consumption in Denmark as the threshold.

8.3 Charity Fundraising

While the main focus of this paper is somewhat technical and abstract, the methods used can readily be applied to the domain of charity fundraising. This section provides an illustration of how the estimation of preferences could increase charitable giving, if projects rather than people are assumed to be the *others*.

Imagine a charity. Within the charity there are four *projects*: Water, Education, Shelter and Medication. The aim of the charity is to raise money to enable the projects to be funded. In order to do so, there are alternative fundraising *campaigns* which can be undertaken, which encourage people to donate. Each campaign advertises alternate bundles of the projects. There are sixteen possible campaigns which can be delivered to potential donors:

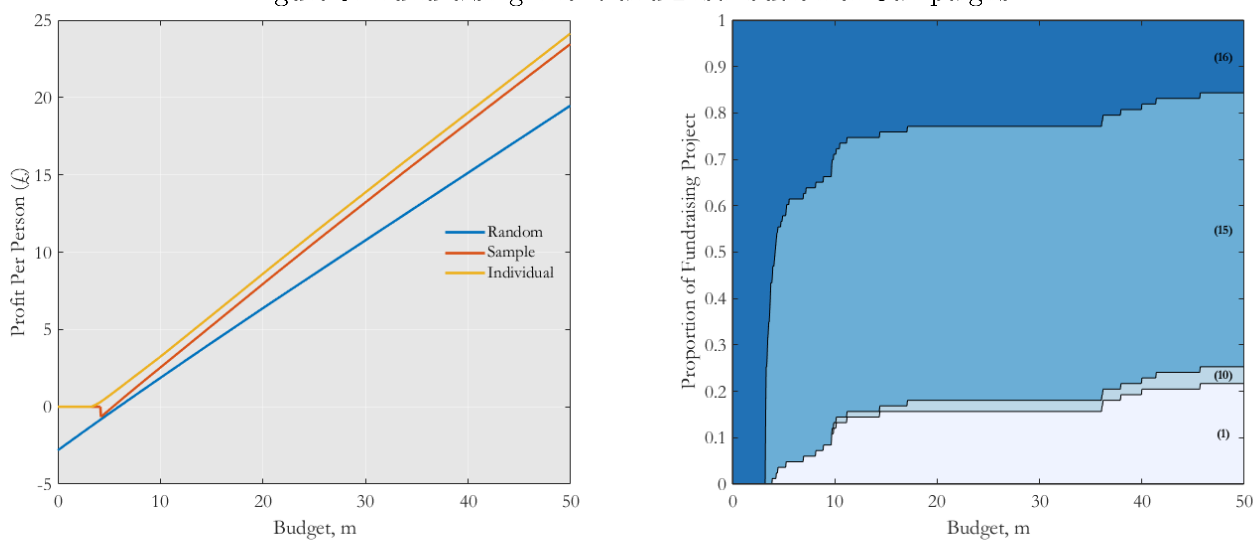
1. Water
2. Education
3. Shelter
4. Medication
5. Water, Educ
6. Water, Shelter
7. Water, Medication
8. Education, Shelter
9. Education, Medication
10. Shelter, Medication
11. Water, Education, Shelter
12. Water, Education, Medication
13. Water, Shelter, Medication
14. Education, Shelter, Medication
15. All Projects
16. Do Nothing

In order to advertise a campaign there is a cost of £3 per person, with the exception of (16) in which no campaign is run. Within each campaign information will be provided about the respective projects. Each project has differential fixed costs which determine the 'cost-effectiveness' of that project; stated as "for every pound given the amount of money going directly to that project is X", where X is 50p for Water, 33.3p for Education, 25p for Shelter and 25p for Medication. The charity's task is then to deliver the campaigns which raise the most amount of money. The following analysis addressed this problem.

Donors are assumed to have a budget, m, which they can distribute between consumption (which is entirely cost-effective) and donations to particular charity projects. Using estimated preferences parameters for each of the 83 participants of the experiment (according to each individual's 'type') predictions of how they would optimally allocate between themselves (the 'donors') and each project, within a given campaign, can be made. The advertising costs of the campaigns can be deducted and the average profit per person calculated.

The left panel of Figure 9 shows the calculation of average profit, for varying levels of m, for three alternative methods of choosing the fundraising campaign. The *random* method denotes the profits that would be made if the charity had no information about the preferences of the donors. Here, as there is no information, charities would randomly choose a campaign to send. The *sample* method uses the representative agent preferences (from the *standard model* in Table 4) to identify the optimal campaign to advertise, for each m. The *individual* method uses individual-level preference parameters to establish the optimal campaign to advertise to each individual. The results show that the profits from the *individual* method is always greater or equal to the *sample* method, which is turn is always greater or equal to the *random* method. For low values of m the *random*, and even *sample*, method give negative profits, as the low budget means that the advertising costs are not exceeded by the donations. As m increases the *random* method diverges from the *sample* and *individual*, making relatively lower profits.

Figure 9: Fundraising Profit and Distribution of Campaigns



The right panel shows the proportion of each campaign advertised, under the *individual* method, for differing m. The four campaigns selected are: (1) Water; (10) Shelter, Medication; (15) All Projects; and (16) Do Nothing. At very low m the optimal campaign is to do nothing, as the donations do not exceed advertising costs. As m increases campaigns (1), (10) and (15) are sent to particular individuals, when $m = 50$, the campaign with the largest proportion is (15), with 59.04%, next is (1) with 21.69%, followed by (16) with 15.66% with 3.61% being selected for (10). The reasons for the differences lie in individuals preferences. Those for whom (16) is optimal tend to be self-interested, the mean self-interest parameter of the group is 0.976. Individuals who donate most in (1) are all efficiency prioritarians, with a mean inequality aversion parameter of -0.112. The three individuals within (10) are all classed as *andreoni* types, who are (slightly) weighted prioritarians with high values of b (0.965 on average), meaning they consider *total* rather than *average* payoffs to others. Those within (15) tend to be weighted prioritarians, with substantial regard for others; 26 of whom are *standard* types with positive r, 8 are *FKM* types with positive $r1$, 7 are *andreoni* types with low b, and 7 who are *stone-geary* types with a positive τ. Being able to account for individual preferences allows for selection, which in turn allows for an increase in profit per person.

While this section is primarily illustrative, there are a number of extensions which could be conducted to make it more applicable and realistic. The first relates to error. In the analysis above the assumption is that donors act optimally and according to the preferences estimated, however, there could be error in those predictions. By incorporating the error model proposed, then a monte-carlo simulation could be ran, to establish the optimal campaigns to run, given the error made. The second extension relates to the fundraising aim of the charity. Two types of funding are commonly found in charitable giving, *restricted* and *unrestricted* funding. The above assumes that donations are unrestricted, meaning that the charity can allocate resources to any project they need. However, (especially with a move towards the tracking and accountability of individual donations) donors may give restricted funding, meaning that only those projects they give directly to can be allocated that funding. These considerations can be incorporated into the analysis, selecting the optimal set of campaigns to increase the funding of the 'worst-off' charity, rather than maximising the total profits (equality vs efficiency criteria). Finally, the incorporation of the value of acquiring information is important. While it is clear that the *individual* method performs the best it may be more costly to acquire information on individual level preferences. Collecting information at an aggregate level (perhaps one decision problem, rather than 30) could prove to be less costly, but if this information cost exceeds the gains made above the *random* method then it is counter productive to gather such information. By accounting for the value of information the choice of method optimised at different budget levels.

While experiments ran in the laboratory may appear abstract, external parallels do emerge. By utilising the methods proposed and accounting for individual preferences real world charitable giving could perhaps be increased.

9 Conclusion

To conclude, through running a modified N-person dictator game both *between-other* and *self-other* distributional trade-offs have been investigated as the number of players increases. Results have found that, on average, the proportional payoffs given to the self decrease, as the number of others increases, but not to the extent that the proportion of payoffs to each others remains constant. The majority of the sample are shown to have other-regarding preferences (91.6%), where the majority are classed as 'Weighted Prioritarians' (55.3%), with significant proportion classed as either 'Efficiency Prioritarians' (22.4%) or 'Maximin' (18.42%).

The importance of estimating preferences within alternative utility functions has been shown, with intuitive *extended* preference parameters of: *self-other* and *between-other* inequality aversion, *congestion* and *minimum thresholds*; better explaining the behaviour of particular individuals. The importance of incorporating both goodness-of-fit and predictive accuracy has been shown, alongside considerations of 'information criteria'. The *amalgamated* model (the most complex) provided the 'best' fit for the modal group of participants; however, when accounting for predictive accuracy and 'information criteria' it performed 'best' for no individuals in the sample. Splitting the sample into 'types', of the 83 participants, we observe 33 individual's behaviour is best explained by the *standard* model, 13 by the *FKM*, 21 by *andreoni* and 9 by *stone-geary*; with 11 individuals being classed as *egoists*. Values from the *likelihood proportion* reveal the 'best' utility functions, combined with the Dirichlet error model, well fit and predict individual-level behaviour, with only 5 participants with $\iota < 0.5$.

Prosocial behaviour and distributional preferences have been shown to be extremely heterogeneous. Not only do particular preferences within utility functions best explain their certain individual's behaviour, but alternative models best suit different individuals. Varying the number of players to whom participants can give to may complicate modelling decision making, but it is something we regularly do as humans and is, therefore, something worthy of striving to explain.

References

Akaike, Hirotogu (1998). "Information theory and an extension of the maximum likelihood principle". In: *Selected Papers of Hirotugu Akaike*. Springer, pp. 199–213.

Andersen, Steffen et al. (2008). "Eliciting risk and time preferences". In: *Econometrica* 76(3), pp. 583–618.

Andreoni, James (2007). "Giving gifts to groups: How altruism depends on the number of recipients". In: *Journal of Public Economics* 91(9), pp. 1731–1749.

Andreoni, James and John Miller (2002). "Giving according to GARP: An experimental test of the consistency of preferences for altruism". In: *Econometrica* 70(2), pp. 737–753.

Andreoni, James and Charles Sprenger (2012). "Estimating time preferences from convex budgets". In: *The American Economic Review* 102(7), pp. 3333–3356.

Barr, Abigail et al. (2015). "Economic status and acknowledgement of earned entitlement". In: *Journal of Economic Behavior & Organization* 118, pp. 55–68.

Bonacich, Phillip et al. (1976). "Cooperation and group size in the n-person prisoners' dilemma". In: *Journal of Conflict Resolution* 20(4), pp. 687–706.

Cason, Timothy N and Vai-Lam Mui (1997). "A laboratory study of group polarisation in the team dictator game". In: *The Economic Journal* 107(444), pp. 1465–1483.

Charness, Gary and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests". In: *Quarterly Journal of Economics* 117(3), pp. 817–869.

Dirichlet, Peter Gustav Lejeune (1839). "Sur une nouvelle méthode pour la détermination des intégrales multiples". In: *Journal de Mathématiques Pures et Appliquées* 4, pp. 164–168.

Dixit, Avinash K (1990). *Optimization in economic theory.* Oxford University Press on Demand.

Dolbear, F Trenery et al. (1968). "Collusion in oligopoly: an experiment on the effect of numbers and information". In: *The Quarterly Journal of Economics* 82(2), pp. 240–259.

Engelmann, Dirk and Martin Strobel (2004). "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Reply". In: *The American Economic Review* 94(4), pp. 857–869.

Engelmann, Dirk and Martin Strobel (2007). "Preferences over income distributions: Experimental evidence". In: *Public Finance Review* 35(2), pp. 285–310.

Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis (2011). "Relative earnings and giving in a real-effort experiment". In: *The American Economic Review* 101(7), pp. 3330–3348.

Fehr, Ernst, Michael Naef, and Klaus M Schmidt (2006). "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment". In: *The American Economic Review* 96(5), pp. 1912–1917.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits (2007). "Individual preferences for giving". In: *The American Economic Review* 97(5), pp. 1858–1876.

Forsythe, Robert et al. (1994). "Fairness in simple bargaining experiments". In: *Games and Economic Behavior* 6(3), pp. 347–369.

Fouraker, Lawrence and Sidney Siegel (1963). "Bargaining and group decision making". In: *New York/Toronto/London 196o*.

Geary, Roy C (1950). "A note on "A constant-utility index of the cost of living"". In: *The Review of Economic Studies* 18(1), pp. 65–66.

Hannan, Edward J and Barry G Quinn (1979). "The determination of the order of an autoregression". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 41(2), pp. 190–195.

Hoffman, Elizabeth et al. (1994). "Preferences, property rights, and anonymity in bargaining games". In: *Games and Economic Behavior* 7(3), pp. 346–380.

Isaac, R Mark and James M Walker (1988). "Group size effects in public goods provision: The voluntary contributions mechanism". In: *The Quarterly Journal of Economics* 103(1), pp. 179–199.

Isaac, R Mark, James M Walker, and Arlington W Williams (1994). "Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups". In: *Journal of public Economics* 54(1), pp. 1–36.

Karni, Edi, Tim Salmon, and Barry Sopher (2008). "Individual sense of fairness: an experimental study". In: *Experimental Economics* 11(2), pp. 174–189.

Macro, David and Jeroen Weesie (2016). "Inequalities between others do matter: evidence from multiplayer dictator games". In: *Games* 7(2), p. 11.

Marwell, Gerald and David R Schmitt (1972). "Cooperation in a three-person Prisoner's Dilemma." In: *Journal of Personality and Social Psychology* 21(3), p. 376.

Panchanathan, Karthik, Willem E Frankenhuis, and Joan B Silk (2013). "The bystander effect in an N-person dictator game". In: *Organizational Behavior and Human Decision Processes* 120(2), pp. 285–297.

Parfit, Derek (1997). "Equality and priority". In: *Ratio* 10(3), pp. 202–221.

Rawls, John (1999). *A Theory of Justice; Revised Edition.* Cambridge, MA: Belknap.

Robson, Matthew (2017). "Inequality Aversion, Self-Interest and Oneness: A Ugandan Lab-in-the-Field Experiment". In: *DERS Discussion Paper* 17(12).

Schumacher, Heiner et al. (2017). "One, two, many—Insensitivity to group size in games with concentrated benefits and dispersed costs". In: *The Review of Economic Studies* 84(3), pp. 1346–1377.

Schwarz, Gideon et al. (1978). "Estimating the dimension of a model". In: *The annals of statistics* 6(2), pp. 461–464.

Stone, Richard (1954). "Linear expenditure systems and demand analysis: an application to the pattern of British demand". In: *The Economic Journal* 64(255), pp. 511–527.

A Appendices

A.1 Instructions

Instructions

Welcome. Thank you for coming today.

Please **Read These Carefully**.

Everyone Will Receive the **Same** Instructions.

General Instructions

In this experiment you will be making **decisions** about the **distribution of payoffs** between **yourself** and **other participants** in this room. These **payoffs** are in addition to your **turn-up fee of £3**.

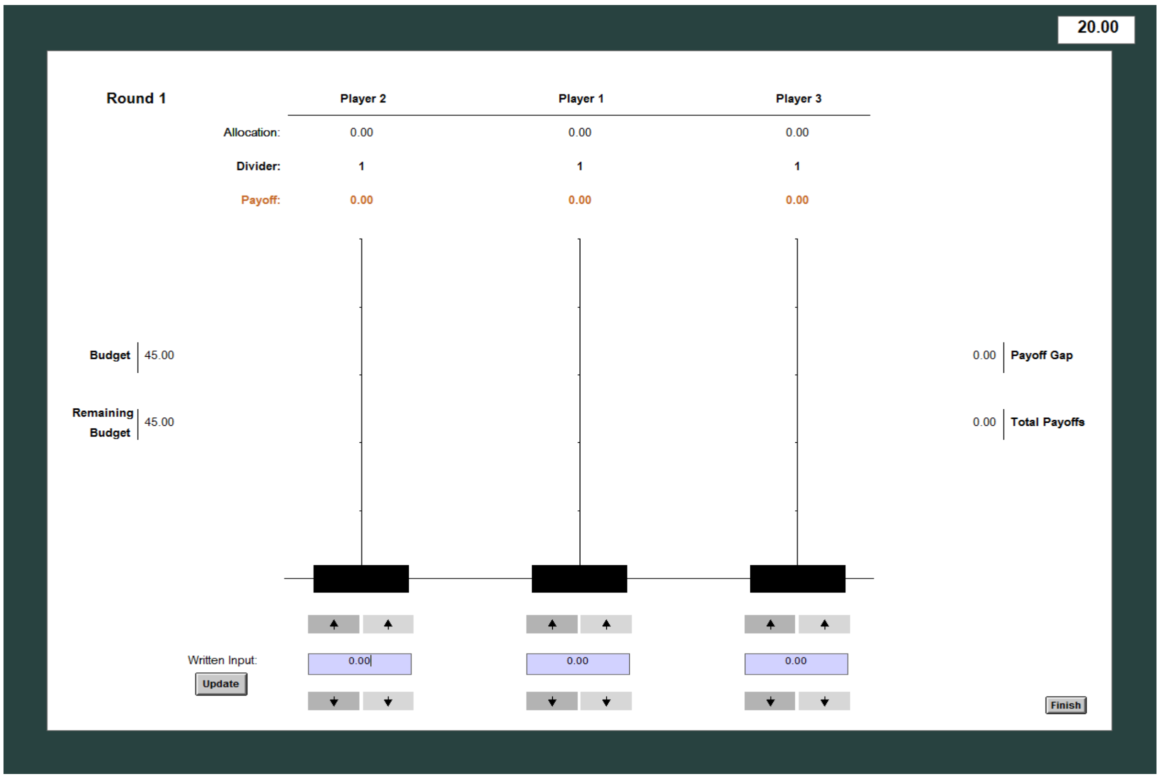
There will be **two** different **stages**, each made up of multiple **rounds**. Your **actual payoff** will be determined from **one randomly selected round**. It is from this **one round** that **all participants** will receive their **payoff**. This means that **every round** has an **equal chance** of determining your **final payoff**, so consider each choice you make **carefully**. **Everyone** will finish at the **same time**, as you will need to wait for **every participant** to finish **each round** before you can move onto the next.

The **individual choices** you make will involve payoffs for **multiple players**. You will make choices which concern the **distribution of payoffs** between those players. In **each round**, you will be **randomly** grouped with some **other** players. Each of you will make your choices **independently** of one another, but only **one** of the player's **individual choices** will be selected, randomly, to provide the payoffs for **all players within that group**. You **will not** know whose decision has been chosen and will receive your payoff, in private, at the end. This means that **every decision** you (and the others) make is **entirely anonymous**. There are **no right or wrong answers**, the decisions you make are entirely up to you and will determine the potential payoffs for you and the others in that group.

The money you have to **allocate** amongst the group will come from a **Budget**. You must decide how to allocate **all** of the money from the **Budget**. For each player in your group, the **Allocation** that you make to them will then be divided by a **Divider** to give their **Payoff** in that round. The **Payoff** can be thought of as the final amount of money that each player gets in that round.

To make your decisions you will be using a **computer interface**, a screenshot of one of the rounds is shown on the next page. Importantly, **you will always be Player 1**, and the other players are **real other participants** in the room.

The **on-screen order** of each player **will vary**. In the example shown here, Player 1 is in the middle.



You will be given a **Budget** to **Allocate** amongst the group. This is shown on the left of the screen, and in this example is £50. You **must** spend the **entire budget** in each round. This means the **Remaining Budget must be zero**. You will be able to make the allocations in three ways. The first is with the **sliders**; you can **drag the sliders** to any allocation that you want. The second is with the **arrow keys**. They allow you to make increases and decreases of 10*p* and 1*p*, respectively. The third is the **written input**; you can click in each of the blue boxes, **type** your desired allocations and **click update**.

Within each of the rounds there will be **different Dividers** for each player. The actual **Payoff** that each player will get will be the **Allocation** you give, divided by the **Divider**. For example, if you give an Allocation of £10, and the Divider is 2, the Payoff will be £5. These **Dividers** are important as they **change every round**, but are **predetermined** and **not dependent** upon your choices.

The **Payoffs** are the final **amount of money** which **will** be given to **each participant**; they will be always be in **pounds** and will be shown by the **height of the orange bars**, the **orange numbers** beside them and the numbers at the **top** of the screen.

Throughout the **rounds**, two elements will change. The first is the **Budget**, so be sure to consider exactly how much the Budget is **before** beginning each decision, as it will vary by a **considerable amount**. The second is the **number of players** in your group. This will change as you go through the experiment. You will see **2, 3, 4, 6** and **12 players** in the groups, throughout various rounds. So remember that **each** of these players is a **real participant** in the room, who will be **anonymous** and **randomly** chosen for each round.

There are **two stages** in this experiment. The first is where you will have **multiple sliders**, one for each Player. The second is where you will have a **single slider** which determines the share of the Budget you choose to give to Player 1, where the other players allocations are equalised.

Remember you are always Player 1. Take note especially in the first stage, as the **order of the players on-screen changes between rounds**.

You will also see the **Total Payoffs** and the **Payoff Gap**. The **Total Payoffs** is the Payoffs of all players added together. The Payoff Gap is the highest Payoff minus the lowest Payoff. Notice how these change when making your decisions. You **must** make a decision in **every** round, and then click **Next** or **Finish** to confirm your decision.

A **minimum time** will be displayed in the **top right corner** in every round, in black. This time **must have elapsed** before you progress to the next round. There will also be a **maximum time**, in red, which will be **double the minimum time**, you **must make** a decision in this time and **click Finish**. **If not**, will receive a **Payoff of zero** for that round and one of the other participants in your group will be the individual whose decisions will count for that round.

After the experiment you will be asked to fill out a **questionnaire**. Your responses from the **questionnaire**, and from the **entire experiment**, will be treated **anonymously**.

After reading these instructions you will go through an **on-screen tutorial**, which will explain how to use the **computer interface** and the exact nature of the experiment. You will then be allowed several **practice rounds** (which **will not** affect your payoff) before making your decisions for **real**.

If you require help at any time, **please raise your hand**.

Please proceed to the On-Screen Tutorial.

A.2 Design and Demographic Differences in Proportional Payoffs

A.2.1 Design

By running separate random effects models for each N, and focusing upon the *multiple* slider treatment, further analysis can be conducted on more specific design effects. Table 10 models PP to Pj, for $N = 2, 3$ and 4, incorporating player specific multipliers, time taken, screen position, the player name and the order of N, alongside the standardised budget. The index j denotes a particular 'other' player, where $j \neq 1, \in N$. Considering the multipliers (π_i), k and l are the 'alternative others', where $k, l \in N$, k is the lowest number that satisfies $k \neq 1, j$, and $l \neq 1, j, k$.

Table 10: Random Effects Model: Proportional Payoff to Pj, Design Effects

	(1) **2 Players**		(2) **3 Players**		(3) **4 Players**	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Multiplier						
- Player 1	-0.1569***	(0.0354)	-0.0908***	(0.0180)	-0.0557***	(0.0123)
- Player j	0.0631***	(0.0198)	0.1140***	(0.0201)	0.1004***	(0.0193)
- Player k			-0.0526***	(0.0184)	-0.0368***	(0.0124)
- Player l					-0.0342***	(0.0106)
Time Finished						
- Average	0.3757***	(0.1299)	0.3273***	(0.0847)	0.1927***	(0.0692)
- Mean Diff: Positive	0.1393***	(0.0454)	0.0238	(0.0226)	0.0315	(0.0251)
- Mean Diff: Negative	-0.0729	(0.0814)	-0.0332	(0.0324)	-0.0144	(0.0226)
Screen Position						
- 2	-0.0000	(0.0124)	0.0083	(0.0075)	0.0024	(0.0063)
- 3			-0.0038	(0.0059)	0.0027	(0.0042)
- 4					-0.0014	(0.0056)
Player Name						
- Player 3			0.0015	(0.0037)	-0.0003	(0.0027)
- Player 4					-0.0023	(0.0023)
N Order						
- Second	0.0598	(0.0419)	-0.0404	(0.0306)	-0.0214	(0.0253)
- Third	-0.0224	(0.0478)	-0.0156	(0.0314)	0.0035	(0.0221)
Standardised N Budget	0.0041	(0.0126)	0.0016	(0.0069)	-0.0022	(0.0073)
Constant	0.2288***	(0.0605)	0.1312***	(0.0383)	0.1056***	(0.0301)
N	83		83		83	
Observations	818		1576		2379	
R-squared	0.1452		0.1496		0.1228	
Between-Subject Variance	0.1662		0.1057		0.0840	
Within-Subject Variance	0.1588		0.1117		0.0880	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results show that as π_j increases the PP to Pj increases, and conversely as $\pi_{i \neq j}$ increases the PP to Pj decreases. The average time participants took to finalise their decision is shown to be strongly positively correlated with the PP to Pj. However, this correlation is perhaps one of reverse causality. The time it takes to allocate all to the self, is much less than ensuring payoffs are distributed equally. The mean difference variables, are constructed to identify within-subject timing differences. For the decisions which participants took longer than their individual average time (within each N) participants give more to others, while when they take less than their average the give less. This relationship is, however, only significant for positive differences in the two player treatment. The effect of the screen position, on giving, is also tested. The base case, is where the screen position is on the left (1), while increasing numbers denote a position further to the right. Results show that there are no significant differences for screen position within either N, for any position. Similar results hold for the name of the 'other' player. The order in which participants made decisions for each N was also varied, the dummy variables show if the order of the N (of that model) was either second or third, with the base case being first. Results show there are no significant order effects. Within N, the budget (which is standardised within each N) has no effect of giving.

A.2.2 Demographics

Alongside the experiment a questionnaire was conducted to establish the demographic characteristics of the participants, alongside questions regarding 'oneness', political persuasion, altruism and beliefs of others giving. The demographic composition of the sample is shown in Table 11. Further to this results from simple random effects models are shown in Table 12. The models include design control variables (number of players, single slider dummy, relative multiplier for P1, standardised budget within each N and round number) and a 'demographic' variable of interest, in order to determine if such variables explain the PP to P1 (over all 45 rounds). A separate model is ran for each, and the resulting coefficient for the variable of interest, alongside standard errors, number of participants and R^2. These are run as separate models for two reasons. The first, missing data. For particular questions a significant proportion of the sample did not answer (in particular parental income and degree subject). The second, is multicollinearity between particular variables, in particular the final four variables concerning altruism and beliefs of the payoffs others gave. As a result a simple modelling approach has been taken, allowing for the comparison of coefficients, significance and model fit; while being wary that these results are prone to omitted variable bias.

Results show that neither age, being an undergraduate, studying science, being of British or Asian nationality, having parents with higher incomes or education, being more right wing have a significant effect on giving to the self. Surprisingly, neither does hypothetical donations, nor willingness to donate to good causes. Females are somewhat more generous, as are art/humanities students, those who are religious, who come from a larger family or are more liberal. While having more friends in the session does not increase giving, a greater 'oneness' (the closeness of connection to others) to the group does. The hypothetical slider questions on "how do you believe the others in this session distributed payoffs" and "what do you believe is a fair distribution of payoffs between yourself and one other" are highly correlated with giving. Those regressors which give the most explanatory power are the fair payoffs, beliefs of payoffs given by others and the 'oneness' to others in the group.

Table 11: Sample Characteristics

	No.	%
Gender		
Male	38	47.5%
Female	42	52.5%
Total	80	100.0%
Age		
18-21	24	29.6%
22-25	30	37.0%
26-29	13	16.0%
30+	14	17.3%
Total	81	100.0%
Nationality		
Asian	37	46.3%
British	32	40.0%
European	8	10.0%
Other International	3	3.8%
Total	80	100.0%

	No.	%
Subject		
Arts and Humanities	18	27.3%
Science	22	33.3%
Social Science	26	39.4%
Total	66	100.0%
Degree Level		
Postgraduate	39	49.4%
Undergraduate	40	50.6%
Total	79	100.0%
Religion		
Agnostic/Atheist	10	14.3%
Christian	10	14.3%
Muslim	12	17.1%
None	32	45.7%
Other	6	8.6%
Total	70	100.0%

Table 12: Random Effects Model: Proportional Payoff to P1, Demographic Effects

	(1) **PP to P1** Coef.	Std. err.	N	R2
Age	0.0022	(0.0045)	81	0.0757
Gender	-0.1043**	(0.0520)	80	0.1034
Undergraduate Dummy	0.0447	(0.0519)	79	0.0891
Arts/Humanities	-0.1110*	(0.0607)	66	0.0897
Science	-0.0324	(0.0660)	66	0.0650
Social Science	0.1223**	(0.0583)	66	0.1008
British	0.0786	(0.0540)	80	0.0920
Asian	-0.0722	(0.0520)	80	0.0890
Religious	-0.1149**	(0.0562)	70	0.1034
Parental Income	-0.0001	(0.0232)	49	0.0718
Parental Education	0.0533	(0.0782)	70	0.0874
Family Size	-0.0236**	(0.0116)	80	0.0923
Oneness - Group	-0.0449***	(0.0137)	83	0.1275
Friends in Session	-0.0082	(0.0184)	80	0.0779
Liberal - Authoritarian	-0.0369***	(0.0130)	80	0.1128
Left - Right	-0.0098	(0.0144)	80	0.0764
Donate	-0.0001	(0.0001)	82	0.0855
Good Cause	-0.0152	(0.0095)	83	0.0986
Fair Payoffs	0.8202***	(0.0844)	82	0.3660
Belief Others Payoffs	0.5937***	(0.1347)	82	0.1455
Controls	YES			

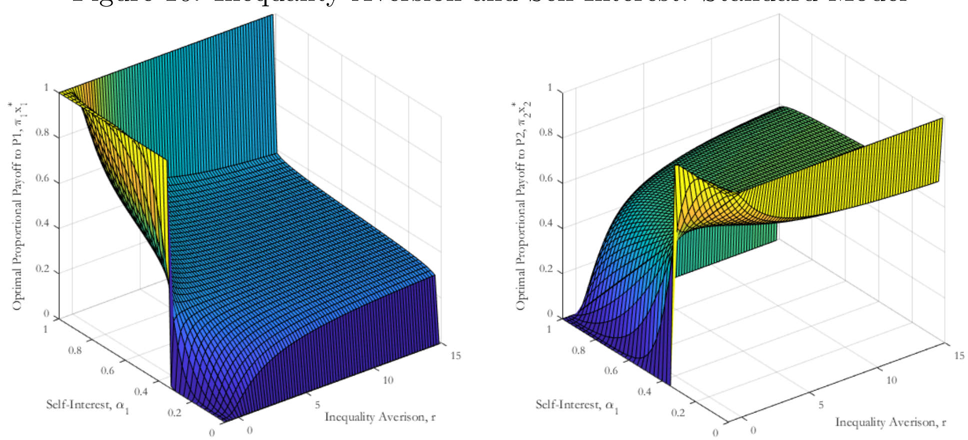
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3 Graphical Intuition of Utility Functions

In addition to the formal notation in Section 3 the following surface plots illustrate the graphical intuition behind the preference parameters. Each figure plots the optimal proportional payoffs for each player i ($\pi_i x_i^*$) for particular preference sets and numbers of recipients.

Figure 10 plots $\pi_1 x_1^*$ in the left panel and $\pi_2 x_2^*$ for different parameter values of self-interest, α_1, and inequality aversion, r. For simplicity the design parameters are set as: $N = 2$, $\pi_1 = 1$ and $\pi_2 = 0.5$. In general as α_1 increases the payoffs to Player 1 increase, while those to Player 2 decrease. Indeed at the extremes, when $\alpha_1 = 1$ then $\pi_1 x_1^* = 1$ and when $\alpha_1 = 0$ then $\pi_1 x_1^* = 0$. The extent to which α_1 changes behaviour depends on r. When $r \to -1$ efficiency concerns are important and so optimal payoffs reflect the highest weighted payoffs that can be obtained. As $r \to 0$ preferences approach Cobb-Douglas, where allocations (x_1^*) are directly proportional to α_1. As $r \to \infty$ equality is the primary concern, and so payoffs become more equal.

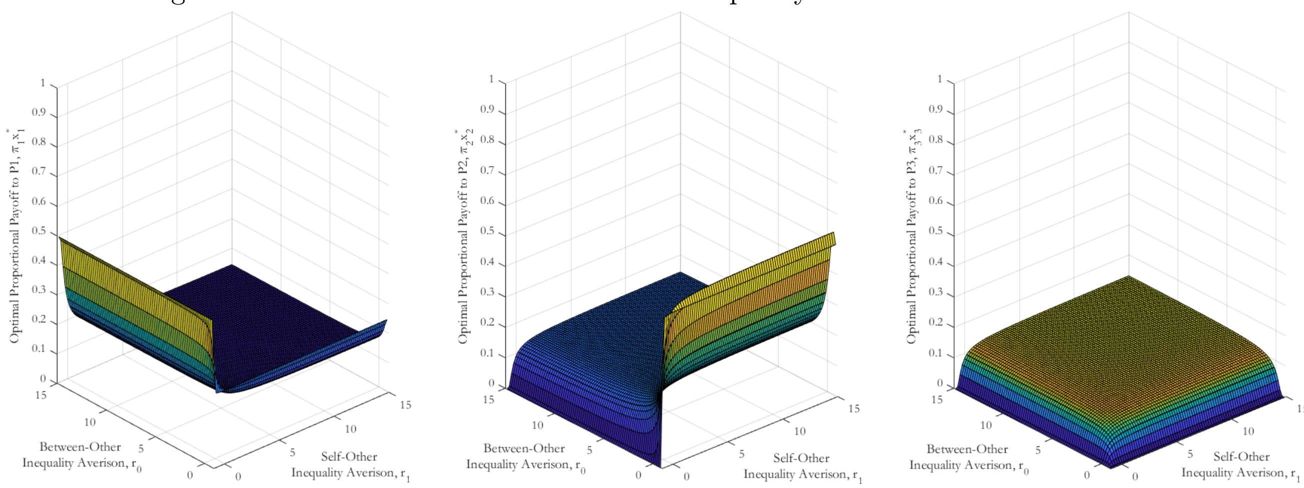
Figure 10: Inequality Aversion and Self-Interest: Standard Model



The *FKM* model allows for the distinction between *self-other* and *between-other* inequality aversion. As the experimental design allows $N > 2$ (and enables participants to distribute between others) this allows for differential behaviour to be observed, and explained by the model. To illustrate this Figure 11 shows the optimal PP to P1, P2 and P3 for varying values of r_1 and r_0; where $\alpha_1 = 0.5$ and the design parameters are: $N = 3$, $\pi_1 = 0.5$, $\pi_2 = 1$ and $\pi_3 = 0.25$. The variation in π_1 allows for differing behaviour to be predicted. As $r_1 \to \infty$ the payoffs are equally distributed between the self and others, and as $r_0 \to \infty$ payoffs are equally distributed between others. When $r_1, r_0 \to 0$ allocations are proportionate to α, so $\pi_1 x_1^* = 0.25$, $\pi_2 x_2^* = 0.25$ and $\pi_3 x_3^* = 0.0625$. If $r_0 \to -1$ the most efficient allocation between-others is preferred, so the share between Player 2 and Player 3 goes entirely to Player 2. If $r_0 > -1$ and $r_1 \to -1$ then all payoffs are allocated to Player 1. At the extreme when $r_1 = r_0 = -1$ individuals are technically indifferent between payoffs to P1 and P2, as $\pi_2 > \pi_3$ and $\alpha_1 = \pi_1/\pi_2 = 0.5$. When $r_1 = r_0$ then behaviour follows that in the *standard* model.

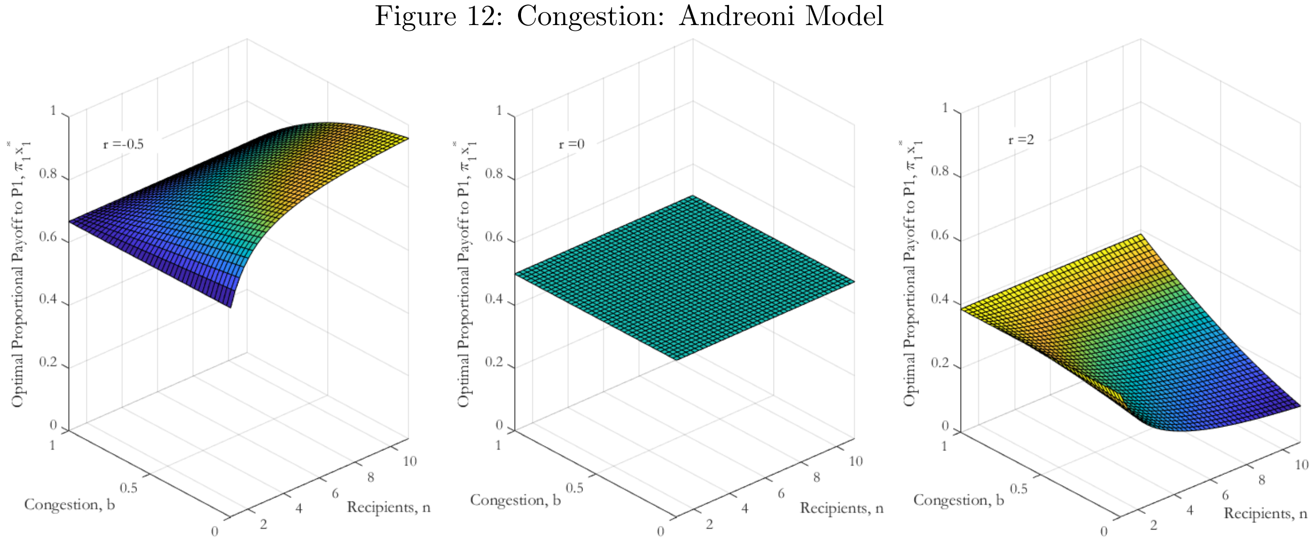
To illustrate how the *andreoni* function models behaviour as n increases Figure 12 plots $\pi_1 x_1^*$ for differing levels of congestion, b, and number of participants n. To simplify, we assume that $\alpha_1 = 0.5$, $\pi_1 = 1$ and $\pi_j = 0.5$ ($\forall j > 1$) but vary r across the three panels, where $r = -0.5$ in the left, $r \to 0$ in the middle and $r = 2$ in the right. If $b = 1$, participants consider the *total* given

Figure 11: Self-Other and Between-Other Inequality Aversion: FKM Model



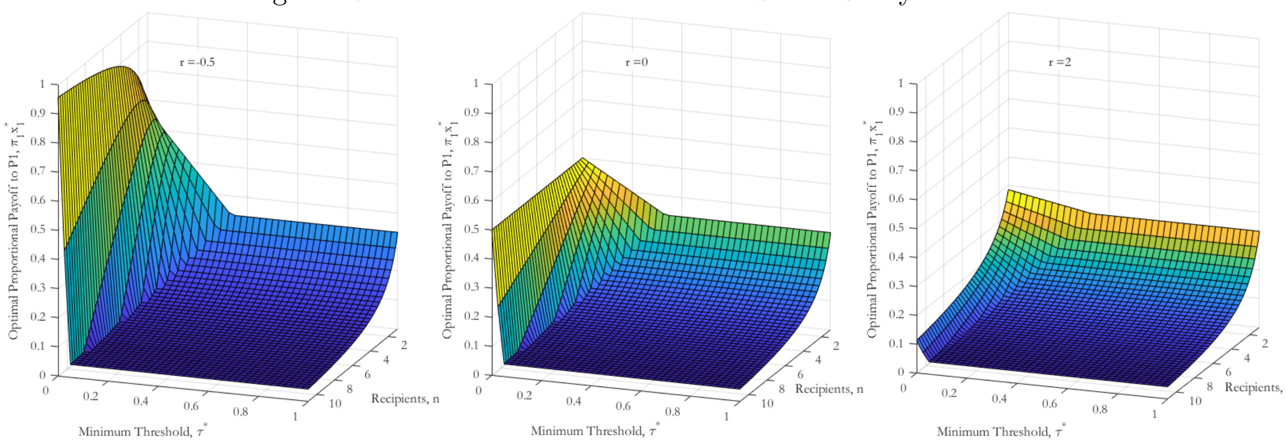
to others, and so as n increases $\pi_1 x_1^*$ remains the same, and therefore $\pi_j x_j^*$ decreases. Indeed, the payoffs to Player 1 are the same for any n if $b = 1$ as when $n = 1$, for any value of b. The different absolute levels between the three panels emerge, as $\pi_1 > \pi_2$; as those prioritising efficiency give more to Player 1, Cobb-Douglas preferences mean that allocations are proportionate to $\alpha_1 = 0.5$ and total payoffs are sacrificed in order to reduce inequality when $r > 0$. As b increases changes in behaviour depend on n and r. If $r < 0$ payoffs to the self increase, if $r \to 0$ then $\pi_1 x_1^*$ remains the same, while if $r > 0$ payoffs to the self decrease. Total payoffs to others actually decrease as n increases when $r < 0$ and $b > 0$ because the weight to each other $((1 - \alpha_1)/n)$ decreases, and so the more 'efficient' choice is to give more to the self. Conversely, when $r > 0$ and $b > 0$, the total payoffs to others increases, as individuals consider the weighted payoff to each other and prefer to weight higher the worst-off.

Figure 12: Congestion: Andreoni Model



The above CES models concern only relative payoffs, and so assume distributions are proportionate to the budget. By incorporating an absolute minimum threshold, τ^*, the effects of a change in the total and average budgets, on behaviour, can be modelled. Figure 13 illustrates how $\pi_1 x_1^*$ changes in relation to τ^* and n (note that the n axis is reversed). An increase of n entails a reduction in the average budget available, and so the effects of differing τ^* can be observed for each n. As above we assume $\alpha_1 = 0.5$, $\pi_1 = 1$ and $\pi_j = 0.5$ ($\forall j > 1$), with r varying across the three panels ($r \approx -0.5, 0, 2$). In each panel for high levels of τ^* the distribution of payoffs are equal between each participant, which ensures $\pi_1 x_1^*$ decreases as n increases as $m = 1$ throughout. When $\tau^* = 0$ the predictions converge to the predictions of the *standard* function. For values of τ^* between the two points, r affects decisions made. The minimum threshold level, in effect, allows for participants to distribute equally to a point, and then distribute according to their other preferences. The scope for this latter distribution depends on the available total and average budget.

Figure 13: Minimum Threshold Levels: Stone-Geary Model



The above illustrates the intuition behind the derivations of the optimal distributions of payoffs, for each of the four models. The *amalgamated* model allows for these *standard* and *extended* behavioural concerns to be combined. This flexibility in modelling enables the extensive heterogeneity in behaviour to be accounted for; when the budget, number of recipients and prices of giving change.

A.4 Stone-Geary Non-Negativity Conditions

Due to the inclusion of τ some optimal allocations may lead to allocations where $\pi_i x_i - \tau_i < 0, \forall i$, which is not feasible. To solve this issue in the main analysis the assumption that $\tau = min(\tau^*, m/\sum_i^N \frac{1}{\pi_i})$ is used. An alternative solution is to make no such assumptions and instead incorporate non-negativity conditions. This does not restrict $\tau_i \leq x_i\pi_i$ explicitly, but provides a set of optimality conditions, within which a subset will ensure $\tau_i \leq x_i\pi_i$. Through this approach, the optimal allocations are as follows:

$$x^*_{i\neq k} = \frac{m + \sum_{j\neq i,k}^{N}\left(\left(\frac{\tau_i}{\pi_j}\left(\frac{\alpha_j\pi_j}{\alpha_i\pi_i}\right)^{\frac{1}{1+r}}\right) - \frac{\tau_j}{\pi_j}\right) - \sum_{k\neq i,j}^{N}\left(\frac{\tau_k}{\pi_k}\right)}{1 + \sum_{j\neq i,k}^{N}\left(\frac{\pi_i}{\pi_j}\left(\frac{\alpha_j\pi_j}{\alpha_i\pi_i}\right)^{\frac{1}{1+r}}\right)}, \qquad x^*_k = \frac{\tau_k}{\pi_k} \tag{18}$$

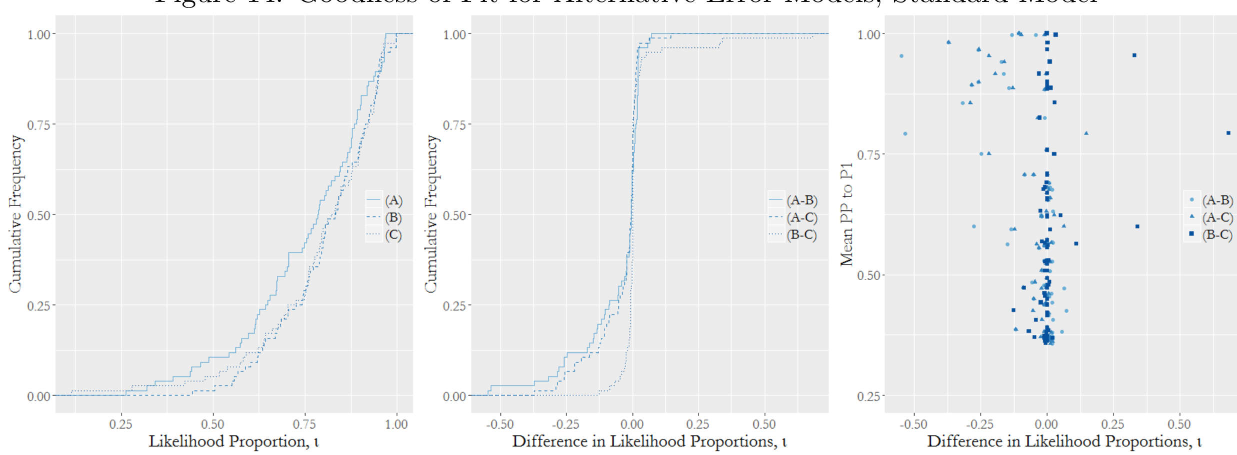
The set of optimality conditions can be concisely written by incorporating k. To solve, we use a solution similar to Lagrange's theorem with non-negative variables (see Dixit (1990) p. 28), which provides equations $\partial L/\partial x_i \leq 0, x_i \geq \tau_i/\pi_i, \forall i$, with complementary slackness, and $\partial L/\partial \lambda = 0$, in order to solve for optimal allocations. In other words, $\forall i$ either $\partial L/\partial x_i = 0$ or $x_i = \tau_i/\pi_i$ (or both). There are $2^N - 1$ combinations of equations, which provide optimal allocations (i.e. for $N = 2$: $[\partial L/\partial x_1 = 0, \partial L/\partial x_2 = 0], [\partial L/\partial x_1 = 0, x_2 = \tau_2/\pi_2], [x_1 = \tau_1/\pi_1, \partial L/\partial x_2 = 0]$). Vector k, then, indexes all those instances where $x_k = \tau_k/\pi_k$. In order to find the optimal, a series of if conditions are formulated (from each $2^N - 1$ combinations of Equation (18)) to enable the optimal to be found, while not violating the above conditions. The ordering of the if statements is, however, crucial as often several optimality condition ensure $\tau_i \leq x_i\pi_i, \forall i$. While this approach is feasible, the additional complexity, computing time needed and issues of ordering meant that it was not used.

A.5 Alternative Error Modelling

Within the specification of the error modelling two main assumptions are made. (1) $E[X_i] = x_i^*$, and (2) $Var(X_i) = \frac{(x_i^*(x_0^* - x_i^*))}{\hat{s}}$. While (1) is not particularly controversial, objections could be made to (2) when considering the nature of the problem, with differing degrees of complexity, κ. The variance of X_i could indeed depend on how many allocation decisions need to be made, independent of x_i^*. As a result three alternative error models, which define s differently, have been used to estimate preference parameters:(A) $\hat{s} = s$, (B) $\hat{s} = s.\kappa$ and (C) $\hat{s} = s.\kappa^{\gamma}$. The three specifications allow for differing variance assumptions to enter into the error modelling. (A) estimates s, with no consideration of the differing κ, while (B) assumes a positive linear relationship between the precision parameter $\hat{s}$ and κ. (C) parameterises this relationship, including an additional parameter for estimation, γ, where $\gamma \in [-1, 1]$, allowing for flexibility which can be captured by γ. The choice of the error model has consequences for the estimation of preference parameters, through its effect on the shape of the probability density functions from which the log-likelihood is calculated.

Figure 14 shows individual-level goodness-of-fit results for the three error models, from estimates within the *standard* model. The left panel shows the distribution of the likelihood proportion, ι, for each error model. The higher ι the better the model is explaining individual behaviour, relative to the uniform distribution, where $\iota < 0.5$ implies that drawing randomly (from a uniform distribution) better explains an individual's behaviour. The distribution shows that (A) performs worse that (B) and (C), while (B) and (C) are closely matched, with the exception of the worst explained, for which (B) performs somewhat better. There are 10.53% of the sample for whom $\iota \leq 0.5$, in (A), 1.32% in (B) and 5.26% in (C).

Figure 14: Goodness-of-Fit for Alternative Error Models; Standard Model



While these distributions show only the aggregate distributions the middle panel shows the distribution of the difference in ι between the models, for each individual. Take the solid line for example, (A-B) shows the difference between ι_A and ι_B, the lower (and negative) the value the better B performs, the higher (and positive) the better A performs. Both (A-B) and (A-C) show that A tends to perform worse, with 61.84% and 63.16% higher ι values for B and C, respectively. The third cdf (B-C) shows a similarity of performance between the two, with C outperforming B for 52.63%. To identify for whom the error models perform better the right panel shows the aforementioned differences plotted against the mean PP to P1. Here we observe that it is the more self-interested individuals for whom A tends to perform worse. B outperforms A to a greater extent than C, for those individuals, however, it also tends to perform worse than A for those who share more equally. C does not tend to have this latter issue.

The reason for the differences in the right panel, can be explained by considering the formula $a_i = x_i^*(\hat{s} - 1)$. Model (A) has an issue when participants have a high degree of self-interest. It models well behaviour for a particular N, as $a_1 > 1$ and $a_{j>1} < 1$, meaning the pdf asymptotes at $x_1 = 1$. However, as $x_{j>1}^*$ may be higher in rounds with lower N, $a_{j>1} > 1$, meaning the pdf becomes uni-modal at an interior allocation, meaning at the bound (where $x_1 = 1$) the likelihood value is very low. Model (B) solves this issue, allowing a_i to vary with κ, however, it does so at the expense of those who are allocating more equally. For them variance is perhaps not decreasing as κ increases, as indeed their decision problem becomes more difficult to distribute equally. Model (C) then allows for the flexibility of estimation, which ensures that the behaviour of the more self-interested is not modelled badly, but that compensating for that does not lead to worse estimates for those who share more equally. An additional parameter does need to be estimated in (C), but due to the above issues and the additional information that γ carries (C) has been chosen for the main analysis.

A.6 Clustering Amalgamated Preferences

Due to the high dimensionality of the preferences in the *amalgamated* utility function visualising and describing the estimated parameters can be difficult. An alternative method of understanding the distribution of preferences is through cluster analysis. A 'mixture' of individual density functions which accurately fit the data can be estimated within a finite mixture model. The multidimensional ellipses, with specific mean and variance in each dimension, capture patterns in the distribution of preferences. Below results are shown from finite mixture model results for r_1, α, r_0, b and τ.[10] Table 13 shows results where three clusters are optimal (allowing *mclust* to search between 0 and 5 clusters), while Table 14 shows results where eight clusters are optimal (allowing *mclust* to search between 0 and 15 clusters).

In Table 13 we observe that the sample is split into three clusters, the largest is Cluster C (49%), followed by Cluster A (30%), with Cluster B as the smallest (21%). Cluster A consists of those who are slightly averse to inequality, with a high level of congestion and some concern for a minimum threshold. This cluster are the most tightly packed, with the lowest variance for all parameters, bar b, in particular with regard to the inequality aversion parameters. Cluster B consists of weakly 'weighted prioritarians', who have zero congestion and, again, and some concern for a minimum threshold. For b variance is very low, but for each other parameter it lies between A and C. Cluster C captures those who are the most averse to inequality and have the largest τ. There is a degree of congestion, more than B, but less than A, while the variance is the largest for all preferences. The mean self-interest remains similar across the clusters, but is the lowest in C. The mean levels and variance of r_1 and r_0 are similar within each cluster.

Table 13: Finite Mixture Model for Amalgamated Preference Parameters: Three Clusters

	Cluster		
	A Mean (Var.)	**B** Mean (Var.)	**C** Mean (Var.)
Ineq. Aversion, r_1	-0.17	1.77	7.03
	(0.07)	(2.76)	(26.31)
Self-Interest, α	0.77	0.77	0.65
	(0.03)	(0.04)	(0.08)
Ineq. Aversion, r_0	-0.26	1.59	7.57
	(0.11)	(3.24)	(28.50)
Congestion, b	0.69	0.00	0.33
	(0.12)	(0.00)	(0.13)
Min. Threshold, τ	1.04	0.84	3.69
	(1.02)	(1.28)	(33.24)
Proportions	0.30	0.21	0.49

In Table 14 an extended number of clusters accounts for a greater extent of the heterogeneity in preferences. As above the clusters are ordered in relation to r_1, with Cluster A capturing a efficiency prioritarian standpoint, while Cluster H encompassed an extreme aversion to inequality. As before r_0 and r_1 appear similar within most clusters, with the exception of Cluster E which captures

[10] To reduce issues of outliers the r_1 and r_0 estimates greater than 15, where capped at a value of 15.

those who have a higher self-other inequality aversion and Cluster H within which between-other inequality aversion is higher. Self-interest varies to a greater extent between clusters, with the lowest α in Cluster F, this is however, mainly due to the large τ which we observe. Congestion is very low in Clusters D and G and high in Cluster A, while the minimum threshold levels are highest in C and F. Interesting differences between similar clusters can be observed. Clusters D and G have similarly low b and τ, but G has higher inequality aversion and lower self-interest. Cluster C and D have similar levels of inequality aversion, but C has much higher congestion, higher τ and lower α. Clusters G and H have similar levels of r_1, but H has much higher α and a higher r_0, while congestion is higher and τ is lower in H.

Table 14: Finite Mixture Model for Amalgamated Preference Parameters: Eight Clusters

	Cluster							
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
	Mean (Var.)	Mean (Var.)	Mean (Var.)	Mean (Var.)	Mean (Var.)	Mean (Var.)	Mean (Var.)	Mean (Var.)
Ineq. Aversion, r_1	-0.29	0.04	1.29	2.10	6.18	8.25	10.01	10.35
	(0.04)	(0.11)	(3.70)	(2.77)	(25.64)	(19.16)	(0.65)	(22.83)
Self-Interest, α	0.68	0.88	0.50	0.72	0.89	0.27	0.45	0.99
	(0.03)	(0.00)	(0.00)	(0.04)	(0.01)	(0.02)	(0.00)	(0.00)
Ineq. Aversion, r_0	-0.25	0.05	1.76	1.81	3.82	9.43	10.47	14.37
	(0.11)	(0.66)	(6.69)	(3.57)	(8.51)	(15.13)	(6.98)	(2.38)
Congestion, b	0.97	0.30	0.66	0.00	0.41	0.28	0.00	0.24
	(0.00)	(0.08)	(0.05)	(0.00)	(0.11)	(0.14)	(0.00)	(0.11)
Min. Threshold, τ	1.03	0.96	3.05	0.81	1.10	12.14	1.49	0.38
	(0.32)	(1.57)	(6.72)	(1.33)	(4.10)	(48.01)	(1.02)	(0.29)
Proportions	0.16	0.19	0.08	0.17	0.14	0.11	0.07	0.09

Through using finite mixture models the complexities of heterogeneous multidimensional preferences can be more easily summarised, and through it interesting differences within the sample observed.

A.7 Mismatches Between Rankings

While the main analysis identifies 'types' of individuals by considering the information criteria for a combined measure of goodness-of-fit and prediction, there are concerns related to the mismatch in the 'ranking' of alternative utility functions. Here a deeper look into the mismatches between rankings based on the differences between Residual Sum of Squares and Log-Likeihood and amongst alternative error models.

A.7.1 Residual Sum of Squares or Log-Likelihood

The Residual Sum of Squares (RSS), identifies the difference between the optimal and observed allocations. Likelihood, identifies the probability that the allocation is observed, given the utility and error model assumed. The former asks how close, the second how probable. Through this alternative criteria differences in which model is considered best will inevitably emerge. In the main analysis the log-likelihood was the metric used, as indeed the estimation procedure was based on maximising the log-likelihood. Here, the differences between the two can be analysed.

Table 15 shows the rankings of utility functions, if the utility functions had been compared using RSS, rather than the log-likelihood values (in Table 7). Comparing Table 15 with Table 7, we observe similar trends. The *amalgamated* model does better within *raw* RSS fit, than prediction, and in RSS compared to IC, while the *standard* model does the opposite. Final IC results for both are somewhat similar, the modal type is *standard*, with the lowest being *amalgamated*. *Standard*, *stone-geary* and *amalgamated* are 'best' for more individuals, while *FKM* and *andreoni* are best for fewer individuals (compared to the log-likelihood IC). When comparing matching within-individuals we observe 51.31%, 59.21% and 59.21% of the sample have matched rankings for GOF, Pred and Both, respectively.

Table 15: Utility Types: Ranked by Residual Sum of Squares and Information Criterion

	Residual Sum of Squares			**Information Criterion**		
	GOF	Predict	Both	GOF	Predict	Both
Standard	6	20	9	39	46	40
FKM	21	18	19	5	10	10
Andreoni	12	14	13	16	12	13
Stone-Geary	12	12	14	11	5	11
Amalgamated	25	11	20	5	3	2

A.7.2 Alternative Error Modelling

In Appendix A.5 three alternatives error models, (A) $\hat{s} = s$, (B) $\hat{s} = s.\kappa$ and (C) $\hat{s} = s.\kappa^{\gamma}$ are discussed. Using preference estimates from each of (A), (B) and (C), similar rankings to those in Table 7 can be conducted. The result of assuming an alternative error model may, lead to different compositions of 'types' in the sample. Table 16 shows the composition of 'types', by using the information criteria, for each alternative error model.

Results show that, in comparison to (C) which is used in the main analysis, there is one more *amalgamated* type in (A), five more in (B). There are lower numbers of *andreoni* types in both

Table 16: Mismatches in Error Models: Ranked by Information Criterion

	(A)	(B)	(C)
Standard	49	42	33
FKM	9	14	13
Andreoni	11	12	21
Stone-Geary	6	3	9
Amalgamated	1	5	0

alternatives, while the number *FKM* types is the lower in (A), and higher in (B). The number of *standard* types is higher in both (A) and (B), while *stone-geary* types are both lower. While these results are sample aggregates, of most interest is how many subjects are classed as the same 'type' in the alternative models. Those of the same 'types' in (A) and (B) are 65.8%, with 55.3% between (A) and (C) and 59.2% between (B) and (C). There are 69.7% for whom two or more models designate the same 'type', with 44.7% who have the same type in all three.

While it is clear that mismatches between rankings do occur, whether that be the metric used to identify the 'goodness' of the model or from differing estimates according to alternate error models, one main result remains. There is still heterogeneity in which models are 'best'. In none of the specifications does one particular utility function dominate and best explain all individual's behaviour.