FOR ONLINE PUBLICATION

Groupy and Non-Groupy Behavior:

Deconstructing Bias in Social Preferences

Online Appendix

Rachel Kranton, Matthew Pease, Seth Sanders, and Scott Huettel

March 2018

This appendix contains the following:

1. Further details of the experiment, including screen shots, sample questions from the surveys that preceded the group treatments, and descriptions of the algorithms pairing subjects in the group treatments.

2. Goodness of Fit tests

3. Estimation of Mixing Model with Five Types

4. Distributions of Social Preferences for Republicans and R-Independents.

5. Nine Type Latent Class Model: Panel Data Minimal Group & Political Group

Results of Latent Class Model with Nine Types: Utility Parameters and Proportions of the Subject Pool

Utility Function Parameters	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Average Subject
Beta (MG, In)	0.00917***	0.213***	0.0706***	0.0112***	0.0728***	0.00530	0.0163***	0.0813***	0.0312***	0.0282***
	(0.00250)	(0.0190)	(0.00652)	(0.00284)	(0.00784)	(0.00335)	(0.00495)	(0.0158)	(0.0108)	(0.00124)
Rho (MG, In)	0.0229***	0.0209***	0.0188***	0.0106***	0.00223	0.0133***	-0.00197	-0.0493***	-0.00665	0.00938***
	(0.00155)	(0.00299)	(0.00211)	(0.00148)	(0.00198)	(0.00186)	(0.00224)	(0.00932)	(0.00420)	(0.000553)
Beta (MG, Out)	0.0112***	0.231***	0.0597***	0.0369***	0.0735***	0.0114***	0.00489	0.0536***	0.0228**	0.0270***
	(0.00274)	(0.0273)	(0.00538)	(0.00455)	(0.00792)	(0.00367)	(0.00465)	(0.0116)	(0.00921)	(0.00121)
Rho (MG, Out)	0.0259***	0.0148***	0.0160***	-0.0264***	-0.00704***	0.0122***	-0.00572**	-0.0433***	-0.000872	0.00460***
	(0.00168)	(0.00273)	(0.00182)	(0.00242)	(0.00213)	(0.00182)	(0.00230)	(0.00740)	(0.00408)	(0.000496)
Beta (MG, In)-	0.00106	-0.00518	-0.0222***	-0.00317	0.0629***	0.00838*	-0.0184***	-0.0387**	-0.0530***	-0.00222
Beta (POL, In)	(0.00360)	(0.0262)	(0.00770)	(0.00406)	(0.0166)	(0.00480)	(0.00664)	(0.0179)	(0.0170)	(0.00171)
Rho (MG, In) -	0.00124	0.00101	0.000584	0.00319	-0.00797**	-0.00278	-0.00182	0.0304***	0.0409***	0.00104
Rho (POL, In)	(0.00220)	(0.00426)	(0.00278)	(0.00216)	(0.00318)	(0.00260)	(0.00313)	(0.0100)	(0.00940)	(0.000777)
Beta (MG, Out) -	0.00111	-0.0878***	0.0281***	0.00732	0.0354**	0.0118**	-0.000522	0.00103	-0.0386***	0.00397**
Beta (POL, Out)	(0.00370)	(0.0294)	(0.00894)	(0.00681)	(0.0149)	(0.00535)	(0.00644)	(0.0161)	(0.0137)	(0.00177)
Rho (MG, Out)-	-0.00406*	-0.0121***	-0.00874***	-0.00475	-0.0109***	-0.0161***	0.00263	0.00473	0.0198***	-0.00388***
Rho (POL, Out)	(0.00227)	(0.00324)	(0.00244)	(0.00373)	(0.00335)	(0.00244)	(0.00318)	(0.00990)	(0.00674)	(0.000698)
Fraction	0.227***	0.196***	0.172***	0.121***	0.107***	0.0852***	0.0425**	0.0355**	0.0142	1
	(0.0354)	(0.0344)	(0.0323)	(0.0274)	(0.0269)	(0.0237)	(0.0170)	(0.0156)	(0.00996)	-
Observations	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0

Predicted Favoritism Levels for each Type

Group Treatment/Designation	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Average Subject
MG	-1.415	2.096**	0.555	36.65***	7.020***	2.017	4.513	1.683	-6.642	5.137***
	(1.191)	(0.955)	(1.521)	(1.918)	(2.071)	(2.618)	(4.221)	(1.985)	(6.717)	(0.788)
POL	1.692	5.129***	12.57***	41.71***	7.082***	18.78***	-0.399	7.547***	6.892	11.31***
	(1.242)	(1.143)	(1.503)	(1.774)	(1.618)	(2.778)	(4.302)	(2.729)	(4.438)	(0.773)
POL minus MG: Absolute	3.107*	3.033**	12.02***	5.055*	0.0617	16.76***	-4.911	5.863*	13.53*	6.173***
	(1.740)	(1.407)	(2.091)	(2.612)	(2.638)	(3.820)	(6.019)	(3.374)	(8.051)	(1.104)
POL-MG Percent	4.5%	4.4%	17.4%	7.3%	0.1%	24.2%	7.1%	8.5%	19.5%	8.9%
Fraction of Subjects	0.227	0.196	0.172	0.121	0.107	0.085	0.043	0.036	0.014	1

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Prevalence of Subjects of each Type by Order of Group Treatment

Order	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9
POL FIRST Fraction	0.193***	0.103**	0.140***	0.210***	0.175***	0.123***	0.0392*	0.0175	
	(0.0522)	(0.0406)	(0.0464)	(0.0538)	(0.0508)	(0.0438)	(0.0237)	(0.0173)	
MG FIRST Fraction	0.253***	0.270***	0.184***	0.0602**	0.0640**	0.0606**	0.0452**	0.0482**	0.0151
	(0.0478)	(0.0506)	(0.0431)	(0.0261)	(0.0292)	(0.0263)	(0.0217)	(0.0235)	(0.0113)
Difference in Fraction	-0.0600	-0.167***	-0.0441	0.149**	0.111*	0.0622	-0.00597	-0.0307	
	(0.0708)	(0.0647)	(0.0633)	(0.0598)	(0.0583)	(0.0511)	(0.0298)	(0.0292)	
Fraction of Subject Pool	0.227	0.196	0.172	0.121	0.107	0.085	0.043	0.036	0.014

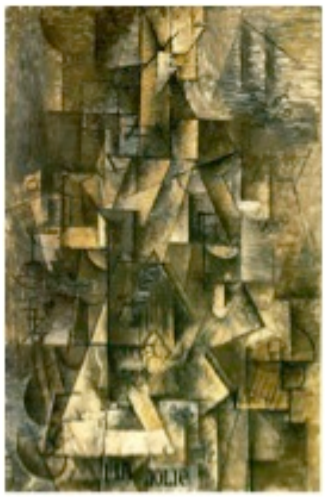
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

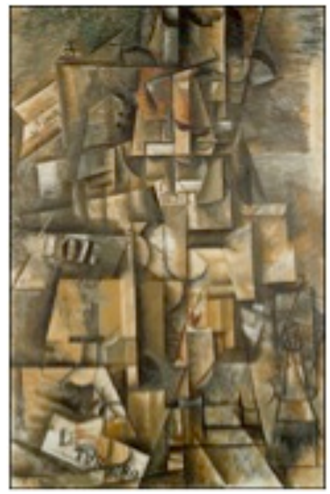
1. Further details of the experiment.

Examples of questions used for the Minimal Group Treatment survey:

Question 4:

Which painting do you prefer?





Question 8:

Which line of poetry do you prefer?

You friendly boatmen and mechanics! You roughs!

You twain! And all processions moving along the streets!

Sample questions on subjects' political affiliation for the Political Treatment survey:

1. Do you consider yourself a(n):

1 DEMOCRAT

2 REPUBLICAN

3 INDEPENDENT

4 NONE OF THE ABOVE

1(a). Are you a strong or moderate DEMOCRAT?

1 STRONG

2 MODERATE

1(a). Do you consider yourself closer to the::

1 DEMOCRATIC PARTY

2 REPUBLICAN PARTY

Pairings in the Minimal Group Treatment: Taken from a bank of other participant's responses, the Own Group Member was selected as the one that answered similarly on the highest number of questions as the subject while the Other Group Member answered most dissimilarly on survey questions. Specifically, an ordered list of subjects who have previously answered the survey is generated. Given a subject *i*, for the Own Group Member, the algorithm goes down this list, counting for each subject the number of similar answers to survey questions. If the second person on the list has a higher number of similar questions, then s/he replaces the first subject as the subject with the highest similarity rating. If a subject further down the list has a greater number of similar answers, s/he then replaces the current most similar subject. Ties are broken in favor of the ordered list. Then, from among the three categories of questions (poetry, painting, and image), whichever category with the highest number of similar responses is selected. Subject *i* is then given the following information: "Overall, the OWN GROUP MEMBER answered [the number of similarly answered questions, overall] survey questions with the same response as you" and "This participant preferred the same [chosen as most similar category, category name] as you on [number of similarly answered questions, in this category] out of 7 questions." A parallel procedure searched for the participant in the pool with the most dissimilar answers to select the Other Group Member and present the corresponding information.

Pairings in the Political Group Treatment: Taken from a bank of other participant's responses, the subjects were divided into two groups. Democrats and subjects answering they were "closer" to Democrats were assigned to the Democrat group. Similarly, Republicans and subjects "closer" to the Republican party were assigned to the Republican group. For Democrats and D-Independents, therefore, the Own Group was the Democrat group, and the Other Group was the Republican group. The subjects were given the following information: "Your OWN GROUP answered similarly on political survey questions," and "The OTHER GROUP answered differently on political survey questions." For the Own Group members selected to be the recipient (by an algorithm that identified a subject that answered similarly on at least one of the five political questions), subjects were presented with the statement "This participant identifies with the [Democrat/Republican] party and subjects were also told the question on which the subject and Own Group Member answered similarly. If the subject and Own Group Member answered several questions similarly, preference was given, in order, to - the abortion, gay marriage, Arizona immigration law, Bush tax cut, and government size questions. Parallel information was given for the Other Group member selected, except the algorithm searched for subjects in the Other Group who had answered dissimilarly on at least one political position question.

Code available upon request.

2. Goodness of Fit

We calculate goodness of fit by considering the fraction of the time the model correctly predicts subjects' actual choices across the 26 matrices and 141 individuals. We conduct the goodness of fit tests for the non-group condition.

Specifically, given the estimated parameters of the utility function, we calculate the utility for the top row and bottom row in each matrix, and assign the *correct choice* as the choice that yields higher utility. We then look at the choice actually made and score the model 1 if the choice made is equal correct choice and 0 otherwise. That is we define

$$G_{im}\left(\hat{\theta}\right)=\begin{cases}1 if\left(\left[u_{im}^{top}\left(\hat{\theta}\right)<u_{im}^{bottom}\left(\hat{\theta}\right)\right]\&\, d_{im}=1\right) or\left(\left[u_{im}^{top}\left(\hat{\theta}\right)\geq u_{im}^{bottom}\left(\hat{\theta}\right)\right]\&\, d_{im}=0\right)\\ 0, otherwise\end{cases} \tag{A1}$$

where θ are the estimated parameters and $d_{im}=1$ indicates person *i* chose the bottom on matrix *m*.

Averaging over all choices of all individuals gives us the fraction of time the model correctly predicts choices. That is, our goodness of fit statistic is

$$\bar{G}\left(\hat{\theta}\right)=\frac{1}{141\times 26}\sum_{i=1}^{141}\sum_{m=1}^{26}G_{im}\left(\hat{\theta}\right). \tag{A2}$$

For some subgroup of people, for example individuals classified as type *t* under the model that allows unobserved heterogeneity, we can classify the goodness of fit of a model for this subtype as

$$\bar{G}_r\left(\hat{\theta}\right)=\frac{1}{N_t\times 26}\sum_{i=1}^{Nt}\sum_{m=1}^{26}G_{im}\left(\hat{\theta}\right) \tag{A3}$$

where person 1 through N_t is in type *t*.

Column (1) of the table below presents this statistic for the Charness and Rabin (CR) model that has one set of parameters for all individuals. On average the model predicts 72.4% of choices correctly. As a point of comparison, the goodness of fit would be 50% for a model in which individuals randomly chose the top or the bottom (as the bottom and top choice were also randomized). The confidence interval (whose construction is discussed below) suggests that the CR model fits much better than a model with random choice – the 95% confidence interval is a goodness of fit of 71.2% to 73.2% - a sound rejection of the random choice model in favor of the CR model. It is also clear that the model fits better for some types than for others. The model fits better for Selfish and Total Income Maximizing types (1 & 2) and more poorly for Inequity Averse and Dominance Seeking types (3 & 4). The last row of column (1) presents the 95% confidence interval around the goodness-of-fit statistics. To calculate this statistic we follow the procedure outlined by Woutersen and Ham (2013).

Goodness of Fit, Confidence Intervals, and Test Statistics

NON GROUP CONDITION

	One Type	Four Types			
		Model using Posteriors		Model using Classification	
	(1)	(2a)	(1) v (2a)	(2b)	(2a) v (2b)
Type					
1	79.5%	91.8%	***	91.8%	
2	77.9%	80.9%	***	80.7%	**
3	62.5%	76.2%	***	75.7%	***
4	62.6%	91.8%	***	91.8%	
Total	72.4%	82.6%	***	82.4%	**
95% CI	(71.2%,73.2%)	(82.2%,83.4%)		(81.9%,82.6%)	

Notes: ** significant a 5% level; *** significant at 1% level

Having estimated the model, a consistent estimate of the asymptotically distribution is that it is asymptotically normal with a mean equal to the estimated parameters and a covariance equal to the estimated covariance parameters. Given this we draw from this multivariate normal distribution and assess whether the draw lies within the 95% confidence interval of the parameter vector. To assess this we conduct a simple chi-square test. The test statistic, $S_d = \theta_d \hat{\Sigma}^{-1} \theta_d^{'}$ is distributed $X^2(k)$ where k is the number of parameters in θ, θ_d is one draw from the normal $N(\hat{\theta}, \hat{\Sigma})$ where $\hat{\theta}$ and $\hat{\Sigma}$ are the estimated values of θ and Σ respectively. All values of θ_d where $S_d < X^2_{.95}(k)$ are used to calculate a series of goodness of fit statistics where values of θ_d where $S_d >= X^2_{.95}(k)$ are rejected as draws of θ inside the 95% confidence interval and hence goodness of fit statistics are not calculated on these values. The goodness of fit for that draw is calculated by simply replacing $\hat{\theta}$ with θ_d in Equation A1 and A2. The Confidence intervals presented in the last row of the table then represent the minimum and maximum values of the goodness of fit test statistics calculated for all of the non-rejected draws of θ_d. What is clear from the table is that there is a relatively tight confidence interval around our estimate of the CR model fitting the data 72.4% of the time.

Columns (2a) and (2b) recalculate the goodness of fit statistics for two versions of our model that allow there to be heterogeneity in parameters by type. Column (2a) calculates the model when the expected utility for choosing the top and the bottom is calculated integrating the expected utility given *i* is of type *t* over the posterior probability of and individual being of each type. That is expected utility in Equation A1 is calculated as $u_{im}^{top}(\hat{\theta}_t) = \sum_{t=1}^{4} u_{im}^{top}(\hat{\theta}_t) \times p_{it}$ where $\hat{\theta}_t$ is now the parameters for type *t* and p_{it} is the posterior probability that person *i* is of type *t* (the utility of the bottom is calculated similarly). Column (2b) expected utility is calculated after assign each individual to the type where their posterior probability of group assignment is highest.

Several issues become clear. First the model that allows heterogeneity fits the data better than the CR model. We estimate that this model fits the data 82.6% (82.4%) of the time. Second, the model improves for all types but especially improves where the CR model fits poorly. The CR model represents social welfare maximizing types well; hence our model only modestly improves on the fit for this group. But the CR model fits dominance seeking individuals very poorly; this is where our model improves fit dramatically. Finally, it does not appear that the model which uses posterior probabilities to calculate expected utilities fits better in a substantive way than the model that classifies individuals by their maximum posterior probability of group membership.

To formally test these propositions about goodness of fit, we use a modified version of Woutersen and Ham (2013). When we test two models against each other we draw from the joint distributions of the parameters of both models. Under each model we calculate the goodness of fit under a specific draw of the joint distribution of θ_d (that meets the chi-square criterion discussed above) and then score which model has a higher goodness of fit. For example, we score under a particular θ_d whether the model that allows for heterogeneity in types fits better than the CR model. We do this for 1000 accepted draws. We then test the null hypothesis that there is no difference in goodness of fit between the CR model and our model in favor of the alternative that our model fits better. We reject the null at the 5% level if we find that the CR model fits better less than 5% of the time (for 50 out of 1000 draws). The column labeled "(1) vs. (2a)" shows that we can reject the null hypothesis that the CR model fits as well as our model overall and for each type and the 1% confidence level.

We repeat the exercise now testing whether the version of our model in column (2a) fits better than the version in (2b). Notice here that there will be a very high covariance between these test statistics. This is because (2a) and (2b) rely on exactly the same parameters for any draw of θ_d. Because of the strong positive covariance between the two test statistics, the precision in the difference is very high. In fact, we find even though there are no meaningful differences between the two versions of our model, we can reject at the 5% level that the two versions perform the same. On purely statistical grounds the model that uses the information on the posterior probability of group membership fits better and is statistically significant. However, as we discussed, the models are not substantively different in their fit.

Woutersen, Tiemen and John C. Ham, "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters," Working Paper, University of Maryland, 2013.

3. Estimation of Mixing Model with Five Types

Throughout the paper we present results for a finite mixture model with four types. This is the minimum number of types needed to describe the four motivations we wish to capture. As is true in all finite mixture models, the parameters estimated are a function of the number of types. Here we consider the quantitative and qualitative effects of estimating five types.

The table below presents estimates from a model that allows for five types and repeats the estimates for the four-type model. These estimates are for the Non-group condition. The parameter estimates for Selfish (type 1), Inequity Averse (type 3) and Dominance Seeking (type 4) types vary very little between the 4-type and 5-type model. The parameter estimates that differ is for Total Income Maximizing (type 2) individuals. In the 4-type model these individuals represent 36.2% of the population. When five types are estimated, a new type 2 set of parameters emerges which consistent with total income maximizing. A new type - type 5 – emerges which can be described as inequity-averse behavior. But this inequity-averse behavior is less inequity-averse than type 3 individuals; type 5 people value own payoff more and inequality in payoff less than type 3 people (but still value both). Effectively, the 5-type model splits the original type 2 individuals into one type that displays stronger income maximizing behavior and a second type that displays weak inequity-averse behavior.

The second table below shows this explicitly. Each individual is classified into a type according to the 4-type and 5-type model. Almost every individual classified as for Selfish (type 1), Inequity Averse (type 3) and Dominance Seeking (type 4) in the 4-type model is classified the same way in the 5-type model. The 5-type model splits the original type 2 people into two types: a new type 2 (with the same but stronger income maximizing behavior) and a new type 5 (a weak form of inequity-averse behavior). A formal test suggests that the 5-type model is a better fit than the 4-type model. The log likelihood is -1631.65 for the 4-type model and -1607.35 for the 5-type model, which has 4 additional parameters. Using a likelihood ratio test, 2 times the difference in the log likelihoods is 48.7. This is distributed chi-squared with 4 degrees of freedom. We can reject the null hypothesis that the models fit equally well in favor of the 5-type model being a better fit at the 1% level.

While the 5-type model is a statistically better fit, qualitatively the data is well described by the 4-type. Since all mixture models are approximations of an underlying continuous distribution of parameter estimates, we present the clearest version in the paper.

Results from Four-Type vs. Five-Type Mixture Model

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

	Four Types Model				Five Type Model				
Utility Function Parameter	Type 1	Type 2	Type 3	Type 4	Type 1	Type 2	Type 3	Type 4	Type 5
Beta	0.152***	0.0655***	0.0312***	0.0367***	0.157***	0.107***	0.0293***	0.0368***	0.0516***
	(0.0134)	(0.00441)	(0.00310)	(0.00980)	(0.0173)	(0.0138)	(0.00331)	(0.00971)	(0.00702)
Rho	-0.00372	-0.0144***	-0.0214***	0.0528***	-0.00495*	-0.0354***	-0.0209***	0.0530***	-0.00527**
	(0.00254)	(0.00157)	(0.00138)	(0.0106)	(0.00280)	(0.00496)	(0.00149)	(0.0102)	(0.00208)
Sigma	0.00489*	0.00544**	-0.00747***	-0.0439***	0.00590*	0.0132***	-0.00924***	-0.0439***	-0.000145
	(0.00287)	(0.00240)	(0.00240)	(0.0169)	(0.00318)	(0.00439)	(0.00295)	(0.0169)	(0.00392)
Observations	3,636	3,636	3,636	3,636	3,636	3,636	3,636	3,636	3,636
Proportions of Type	24.9%	36.2%	34.0%	5.0%	24.3%	19.3%	31.8%	5.0%	19.7%
Category Implied by Parameter	SELFISH	TOTAL INCOME MAX	INEQUITY AVERSE	DOM SEEKING	SELFISH	TOTAL INCOME MAX	INEQUITY AVERSE	DOM SEEKING	INEQUITY AVERSE

Classification of Individuals by Model

	4-Type Model				
5-Type Model	1	2	3	4	Total
1	35	0	0	0	35
2	0	25	2	0	27
3	0	1	45	0	46
4	0	0	0	7	7
5	0	26	0	0	26
Total	35	52	47	7	141

4. Distributions of social preferences for Republicans and R-Independents.

Distribution of Social Preferences, by Condition and Match
REPUBLICANS

PANEL A: NON-TYPE		
	YOU-OTHER	
Type	Freq.	Percent
SELFISH	3	17
TOTAL INCOME	10	56
INEQUITY AVERSE	5	28
DOMINANCE	0	0
Total	18	100

PANEL B: MINIMAL TYPE				
	YOU-OWN		YOU-OTHER	
Type	Freq.	Percent	Freq.	Percent
SELFISH	5	28	6	33
TOTAL INCOME	10	56	3	17
INEQUITY AVERSE	3	17	6	33
DOMINANCE	0	0	3	17
Total	18	100	18	100

PANEL C: POLITICAL TYPE				
	YOU-OWN		YOU-OTHER	
Type	Freq.	Percent	Freq.	Percent
SELFISH	4	22	5	28
TOTAL INCOME	6	33	3	17
INEQUITY AVERSE	8	44	6	33
DOMINANCE	0	0	4	22
Total	18	100	18	100

Distribution of Social Preferences, by Condition and Match
REPUBLICAN-LEANING INDEPENDENTS

PANEL A: NON-TYPE

	YOU-OTHER	
Type	Freq.	Percent
SELFISH	8	38
TOTAL INCOME	4	19
INEQUITY AVERSE	9	43
DOMINANCE	0	0
Total	21	100

PANEL B: MINIMAL TYPE

	YOU-OWN		YOU-OTHER	
Type	Freq.	Percent	Freq.	Percent
SELFISH	6	29	7	33
TOTAL INCOME	3	14	2	10
INEQUITY AVERSE	12	57	8	38
DOMINANCE	0	0	4	19
Total	21	100	21	100

PANEL C: POLITICAL TYPE

	YOU-OWN		YOU-OTHER	
Type	Freq.	Percent	Freq.	Percent
SELFISH	8	38	7	33
TOTAL INCOME	1	5	3	14
INEQUITY AVERSE	12	57	6	29
DOMINANCE	0	0	5	24
Total	21	100	21	100

4. Nine Type Latent Class Model: Panel Data from Minimal Group and Political Group Condition

This section of the online Appendix considers an alternative, simpler utility function and a more complex latent class model that analyzes the panel data from the minimal group and political group conditions of the experiment.

Suppose an individual *i*'s utility is a simple additive function of own and the other's income:

$$U_i(\pi_i, \pi_j) = \beta_i\pi_i + \rho_i\pi_j,$$

where β_i is the value *i* places on own income and ρ_i is the value *i* places on person *j*'s income. In a group setting, the values *i* places on π_i and π_j could depend on the nature of the group division and whether *j* is in *i*'s group or not. We consider utility

$$U_i(\pi_i, \pi_j; g, J) = \beta_i(g, J) \cdot \pi_i + \rho_i(g, J) \cdot \pi_j$$

where $g \in \{MG, POL\}$ again indicates the particular group division and $J \in \{IN, OUT\}$ indicates whether *j* is in *i*'s group or not, yielding eight utility function parameters:

$\beta_i(MG, In)$	$\rho_i(MG, In)$
$\beta_i(MG, Out)$	$\rho_i(MG, Out)$
$\beta_i(POL, In)$	$\rho_i(POL, In)$
$\beta_i(POL, Out)$	$\rho_i(POL, Out)$

We estimate a series of latent class models. Each model posits a number of *types*, *T*, where each particular type *t* has a unique set of the eight function parameters (β_t, ρ_t), and each type *t* is a proportion of the population p_t, where $\Sigma_t\, p_t = 1$. The parameters (β_t, ρ_t) and the proportions are estimated to maximize a likelihood function. To select among these models, we use the BIC and the AIC criteria, which balance the increase in the likelihood function from estimating more types against a penalty for the added model complexity of more parameters.[1]

Formally, we build our analysis as follows. If each individual's type were known, we could estimate a binary choice model for choosing the bottom row in each matrix for individuals of type *t*. Assuming an extreme value distribution for the error terms, the parameters could be estimated for the type *t* individuals by maximizing

$$L(\beta_t, \rho_t) = \prod_{i=1}^{\tau} \prod_{m=1}^{26} \Lambda_{mi}(\beta_t, \rho_t | \pi_i, \pi_j)^{d_{mi}} \left(1 - \Lambda_{mi}(\beta_t, \rho_t | \pi_i, \pi_j)\right)^{1-d_{mi}}$$

where

$$\Lambda_{mi}(\beta_t, \rho_t) \ = \ exp(V_{mi}^{bot} - V_{mi}^{top}) / \left(1 + exp(V_{mi}^{bot} - V_{mi}^{top})\right)$$

and

$$(V_{mi}^{bot} - V_{mi}^{top} | \beta, \rho_t) = \beta_t(\pi_{i,m}^{bot} - \pi_{i,m}^{top}) + \rho_t(\pi_{j,m}^{bot} - \pi_{j,m}^{top}).$$

Since we do not know each individual's type, we condition on an individual being a type and then sum over the distribution of types. That is, for *T* types, we estimate

$$L(\beta, \rho, p) = \prod_{i=1}^{141} \prod_{m=1}^{26} \prod_{t=1}^{T} p_t \Lambda_{mi}(\beta_t, \rho_t | \pi_i, \pi_j)^{d_{mi}} \left(1 - \Lambda_{mi}(\beta_t, \rho_t | \pi_i, \pi_j)\right)^{1-d_{mi}}$$

1. There is little consensus on the best selection criterion (see, e.g., Burnham & Anderson (2004)). We use the BIC and AIC in combination, as discussed below, to determine the model to present in this paper. Other selection techniques, such as estimating a model on a subset of the sample and then testing that model's out-of-sample predictions on another subset of the sample (e.g., Bruhin et. al. (2015)), is not possible with our limited-numbers subject pool.

where $p=(p_1,\ldots,p_T)$ is estimated along with the vectors of utility parameters, (β,ρ), for the T types.[2] Since type numbering is arbitrary we denote type 1 as the type with the highest fraction in the sample; the T[th] type has the lowest fraction in the sample.

We estimate each model of $T=\{1,\ldots,\tau\}$ types and each time check the BIC and AIC criteria. The BIC criterion, values of which are reported in the table below, is essentially tied at eight and nine types, and the AIC criterion suggests more than nine types. Since the nine-type model gives more detail by adding a new type (rather dividing one type into two types) and pulls the selection in the direction of the AIC, we report the nine-type model.[3]

Bayesian Inference Criterion Calculations for Latent Class Models

Features of Latent Class Model					
Number of types	7	8	9	10	11
Number of Parameters	63	72	81	90	99
Log Likelihood	-6194.96	-6055.95	-6015.03	-6003.85	-5980.98
BIC	12993.95	12802.23	12806.69	12870.61	12911.17
Observations	14587	14587	14587	14587	14587

Classifying individual subjects as types, using subject's actual choices in the experiment, shows the estimation indeed well captures behavior. Having estimated the model, it is straightforward to calculate the *posterior* probability that a particular subject i is type t. Under the estimated parameters and given the choices that i actually made, the probability of making those choices if i is type t is

$$\Gamma_t(\beta_t,\rho_t)=\prod_{k=1}^{26}\Lambda_{tk}(\beta_t,\rho_t|\pi_i,\pi_j)^{d_{ki}}\times\left(1-\Lambda_{tk}(\beta_t,\rho_t|\pi_i,\pi_j)\right)^{(1-d_{kt})}$$

Using Bayes' rule with the estimated mixing proportions p_t as priors of being type t, the posterior probability that i is type t, P_t is just

$$P_t(\beta,\rho)=\frac{p_t\Gamma_t(\beta_t,\rho_t)}{\sum_{t=1}^{9}p_t\Gamma_t(\beta_t,\rho_t)}\quad .$$

We then categorize individuals as type t based on their posterior probability of being type t. In particular, we assign i type t if $P_t=max(P_1,\ldots,P_9)$. Of the 141 subjects in our experiment, 128 are assigned to their type with probability at or above 0.99. Only ten subjects have a posterior probability below 0.90. These ten subjects are dispersed among types 2, 3, 4, and 6; hence, no single type absorbs these participants. Finally, all of the subjects categorized as types 7, 8, or 9,

2. To insure that $0\le p_t\le 1$ for all t, the mixing distribution is specified as a logistic function with a constant. That is, T-1 constants $\{\theta_1,\theta_2,\ldots,\theta_{T-1}\}$ are estimated, and the proportion of type 1 is then calculated as $\exp(\theta_1)/(1+(\exp(\theta_1)+\exp(\theta_2)+\ldots+\exp(\theta_{T-1}))$ and similarly for the proportion of each type t.

3. In addition, we checked whether the 8, 9, or 10 types models led to different categorizations of subjects to types (categorization described below). In general, parameter estimates for types were extremely close when subjects were not reclassified as the number of types expanded. One exception is that types 7 and 9 in the nine-type model were combined into a single type in the eight-type model. These two types appear distinct in their behavior, an additional reason we settled on the nine-type model. When we estimated the ten-type model, the additional type was poorly estimated and not behaviorally distinguishable from existing types.

have posterior probabilities of 0.99 or 1, indicating that these types, while small fractions of the population, well portray these subjects' distinct patterns of behavior.

The Tables below present the results of the nine-type model. The parameters for the "average subject," which are estimated from a degenerate model with one type, are presented in the last column. The top half of the table reports the parameter estimates for the minimal group; the bottom half contains the political group parameters by providing the difference between the parameters in the minimal group and the political group for each type, along with tests of whether the differences in the parameters are statistically significant.

An overview of the results shows that Type 1 and Type 4—an estimated 40% of total subjects—each have utility parameters that are statistically the same in the political group and minimal group treatments. The two types, however, are quite different in their weights on outgroup incomes. Type 1 puts similar weights on own income, ingroup income, and outgroup income; Type 4 puts negative weight on outgroup income. The rest of the types have utility function parameters that are statistically different in minimal group and political group, but the implications of the parameters for allocations and the economic significance of these differences is not readily apparent.

To interpret the results in depth, we evaluate both the parameter values and the predicted amounts each type would allocate to ingroup and outgroup recipients in each treatment. The latter are reported in the second table below. From these amounts, we also calculate predicted favoritism levels for each type, also reported in this table.[4]

We proceed by discussing each type in turn, from the most prevalent to the least prevalent. The most prevalent type—estimated 22.7% of subjects—places slightly higher weight on others' income than own income, when the recipient is ingroup or the recipeient is outgroup, in both group treatments.[5] In each group treatment, the predicted favoritism levels are not statistically different than zero, and we cannot reject that the favoritism levels are the same at the 5% level.

Type 2, the next largest type—estimated 19.6% of subjects—favors strongly own income relative to others' income, ingroup or outgroup, in both group treatments.[6] The political group utility parameters are statistically different than those for the minimal group treatment; political group predicted favoritism is statistically positive while minimal group favoritism is statistically zero. This difference is statistically different than zero. However, the difference in the favoritism levels is small and arguably not economically significant, at only 4.4% of the total possible favoritism. Indeed, the difference is just below that of Type 1. If we impose an economic significance criterion of the difference in favoritisms to be at least 5%, this type might be considered to have a bias.

Like Type 2, the third most prevalent type—estimated 17.2% of subjects—places higher weight on own income than other's income, but the magnitudes are smaller. Relative to the minimal group, Type 3's weight in the political group on ingroup income is higher and weight on outgroup income is lower. The predicted favoritism level in political group is both economically and statistically significant, but the minimal group favoritism statistically zero.

---

4. We also calculated: (1) the predicted difference in own income when facing an ingroup vs. an outgroup participant, and (2) the absolute payoff levels that underly these ingroup/outgroup differences; i.e., the predicted payoffs for each type in each match, the predicted payoffs for an ingroup recipient and an outgroup recipient facing each type, as well as predicted expected total surplus for each type. All these payoffs generally track the utility function parameters.
5. These weights are consistent with achieving equal levels of income between self and other subjects. Predicted payoffs own for a decision-maker of this type are the lowest of all types, though predicted aggregated payoffs are the highest among the nine types.
6. Type 2 earns the highest payoffs of all nine types and matches involving Type 8 yield the lowest aggregate payoffs.

The fourth most prevalent type—estimated 12.1% of subjects—weighs ingroup recipient's income much more than outgroup recipient's income, and the utility parameters are statistically the same in the minimal and political group treatment. Consequently, this type has the largest favoritism levels in both minimal group and political group treatment, up to 60% of the maximal difference.

For the fifth type—estimated 10.7% of subjects—the utility parameters are statistically different across group conditions, but translate into positive but small favoritism, about 10% of the total possible, and is statistically the same in the two group conditions.

The sixth type—estimated 8.5% of subjects—is similar to Type 3, but with larger divergence between parameters in minimal group and political group vis-à-vis outgroup recipient. This type, with a difference in favoritism that is the largest among the types at 24.2%.

For the seventh (4.3% of subjects), eighth (3.6% of subjects), and ninth (1.4% of subjects) types, statistical tests have low power, yet we can still see the distinctive behavior of each. Type 7 is the only type whose political vs. minimal group utility parameters only differ for own income and vis-à-vis ingroup; this difference though leads a small difference in favoritism.

Type 8, unlike all other types, puts statistically significant negative weight on recipient's income, whether ingroup, outgroup, political group or minimal group. With the small weight on own income, this type appears to be willing to sacrifice own income in order to reduce the income of others, and more so in the political group. The resulting favoritism difference, however, is not statistically significant.

Finally, Type 9 appears to care only about own income when giving to an ingroup participant and to care about own and other's income when giving to an outgroup participant, and these weights are higher in the political than the minimal group.

For the average subject, neither the utility function nor the favoritism levels match those of any one of the nine types. The favoritism level in minimal group condition is closest in magnitude to that of Type 5 but political group favoritism is closest to that of Type 3 (and for each of these levels, we cannot reject that the respective magnitudes are the same).[7]

In the final table below, we compare the prevalence of types among those who received the minimal group treatment first with the prevalence of types among those who received the political group treatment first. The comparison shows suggestive evidence of a spillover of the more salient, political treatment on the minimal group treatment for one set of subjects. The fraction of subjects of each type are statistically identical except for Type 2, which has fewer political-group-first subjects, and Type 4, which has more political-group-first subjects. Subjects who have the highest levels of favoritism in the political treatment are more likely to have the highest levels of favoritism in the minimal group treatment if they received the political treatment first. The order of group treatments does not appear to matter for the prevalence of any other types.[8]

7. Tests for differences conclude that the average subject's minimal group favoritism level is statistically different than those of Types 1, 2, 3, and 4 at the 0.05 level or below; political group favoritism is statistically different than those of Types 1, 2, 4, and 5 at the 0.05 level or below. Many of the tests relative to Types 6,7,8, and 9 are not sufficiently well powered to detect differences.

8. As another check, we split the sample into minimal-group-first subjects and political-group-first subjects and estimate separately a nine-type latent class model for each. The predicted behaviors of the most prevalent types are qualitatively the same for both subsamples; estimates for the less prevalent types are similar but harder to compare due to small numbers.

Results of Latent Class Model with Nine Types: Utility Parameters and Proportions of the Subject Pool

Utility Function Parameters	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Average Subject
Beta (MG, In)	0.00917***	0.213***	0.0706***	0.0112***	0.0728***	0.00530	0.0163***	0.0813***	0.0312***	0.0282***
	(0.00250)	(0.0190)	(0.00652)	(0.00284)	(0.00784)	(0.00335)	(0.00495)	(0.0158)	(0.0108)	(0.00124)
Rho (MG, In)	0.0229***	0.0209***	0.0188***	0.0106***	0.00223	0.0133***	-0.00197	-0.0493***	-0.00665	0.00938***
	(0.00155)	(0.00299)	(0.00211)	(0.00148)	(0.00198)	(0.00186)	(0.00224)	(0.00932)	(0.00420)	(0.000553)
Beta (MG, Out)	0.0112***	0.231***	0.0597***	0.0369***	0.0735***	0.0114***	0.00489	0.0536***	0.0228**	0.0270***
	(0.00274)	(0.0273)	(0.00538)	(0.00455)	(0.00792)	(0.00367)	(0.00465)	(0.0116)	(0.00921)	(0.00121)
Rho (MG, Out)	0.0259***	0.0148***	0.0160***	-0.0264***	-0.00704***	0.0122***	-0.00572**	-0.0433***	-0.000872	0.00460***
	(0.00168)	(0.00273)	(0.00182)	(0.00242)	(0.00213)	(0.00182)	(0.00230)	(0.00740)	(0.00408)	(0.000496)
Beta (MG, In)-	0.00106	-0.00518	-0.0222***	-0.00317	0.0629***	0.00838*	-0.0184***	-0.0387**	-0.0530***	-0.00222
Beta (POL, In)	(0.00360)	(0.0262)	(0.00770)	(0.00406)	(0.0166)	(0.00480)	(0.00664)	(0.0179)	(0.0170)	(0.00171)
Rho (MG, In) -	0.00124	0.00101	0.000584	0.00319	-0.00797**	-0.00278	-0.00182	0.0304***	0.0409***	0.00104
Rho (POL, In)	(0.00220)	(0.00426)	(0.00278)	(0.00216)	(0.00318)	(0.00260)	(0.00313)	(0.0100)	(0.00940)	(0.000777)
Beta (MG, Out) -	0.00111	-0.0878***	0.0281***	0.00732	0.0354**	0.0118**	-0.000522	0.00103	-0.0386***	0.00397**
Beta (POL, Out)	(0.00370)	(0.0294)	(0.00894)	(0.00681)	(0.0149)	(0.00535)	(0.00644)	(0.0161)	(0.0137)	(0.00177)
Rho (MG, Out)-	-0.00406*	-0.0121***	-0.00874***	-0.00475	-0.0109***	-0.0161***	0.00263	0.00473	0.0198***	-0.00388***
Rho (POL, Out)	(0.00227)	(0.00324)	(0.00244)	(0.00373)	(0.00335)	(0.00244)	(0.00318)	(0.00990)	(0.00674)	(0.000698)
Fraction	0.227***	0.196***	0.172***	0.121***	0.107***	0.0852***	0.0425**	0.0355**	0.0142	1
	(0.0354)	(0.0344)	(0.0323)	(0.0274)	(0.0269)	(0.0237)	(0.0170)	(0.0156)	(0.00996)	-
Observations	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587	14,587

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0$

Predicted Favoritism Levels for each Type

Group Treatment/Designation	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Average Subject
MG	-1.415	2.096**	0.555	36.65***	7.020***	2.017	4.513	1.683	-6.642	5.137***
	(1.191)	(0.955)	(1.521)	(1.918)	(2.071)	(2.618)	(4.221)	(1.985)	(6.717)	(0.788)
POL	1.692	5.129***	12.57***	41.71***	7.082***	18.78***	-0.399	7.547***	6.892	11.31***
	(1.242)	(1.143)	(1.503)	(1.774)	(1.618)	(2.778)	(4.302)	(2.729)	(4.438)	(0.773)
POL minus MG: Absolute	3.107*	3.033**	12.02***	5.055*	0.0617	16.76***	-4.911	5.863*	13.53*	6.173***
	(1.740)	(1.407)	(2.091)	(2.612)	(2.638)	(3.820)	(6.019)	(3.374)	(8.051)	(1.104)
POL-MG Percent	4.5%	4.4%	17.4%	7.3%	0.1%	24.2%	7.1%	8.5%	19.5%	8.9%
Fraction of Subjects	0.227	0.196	0.172	0.121	0.107	0.085	0.043	0.036	0.014	1

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Prevalence of Subjects of each Type by Order of Group Treatment

Order	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9
POL FIRST Fraction	0.193***	0.103**	0.140***	0.210***	0.175***	0.123***	0.0392*	0.0175	
	(0.0522)	(0.0406)	(0.0464)	(0.0538)	(0.0508)	(0.0438)	(0.0237)	(0.0173)	
MG FIRST Fraction	0.253***	0.270***	0.184***	0.0602**	0.0640**	0.0606**	0.0452**	0.0482**	0.0151
	(0.0478)	(0.0506)	(0.0431)	(0.0261)	(0.0292)	(0.0263)	(0.0217)	(0.0235)	(0.0113)
Difference in Fraction	-0.0600	-0.167***	-0.0441	0.149**	0.111*	0.0622	-0.00597	-0.0307	
	(0.0708)	(0.0647)	(0.0633)	(0.0598)	(0.0583)	(0.0511)	(0.0298)	(0.0292)	
Fraction of Subject Pool	0.227	0.196	0.172	0.121	0.107	0.085	0.043	0.036	0.014

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1