FOR ONLINE PUBLICATION

Groupy and Non-Groupy Behavior: Deconstructing Bias in Social Preferences

Online Appendix

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

May 2016

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.

1. Further details of the experiment.

Examples of questions used for the Minimal Group Treatment survey:

Question 4:

Which painting do you prefer?

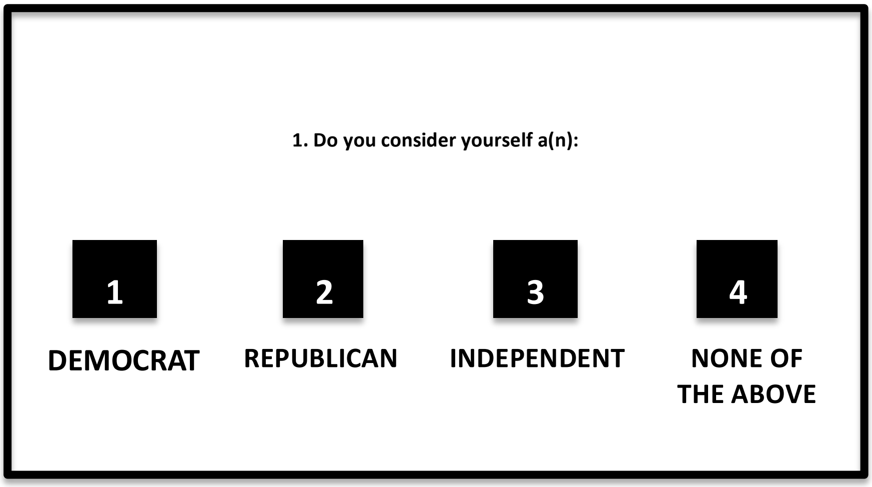
Question 8:

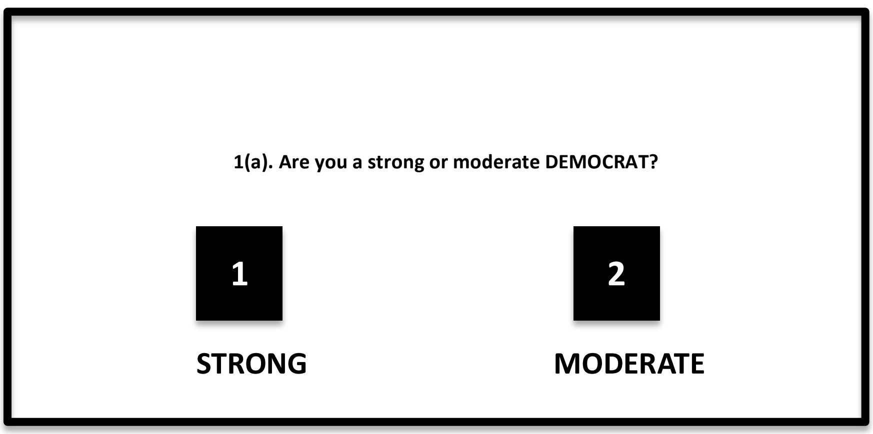
Which line of poetry do you prefer?

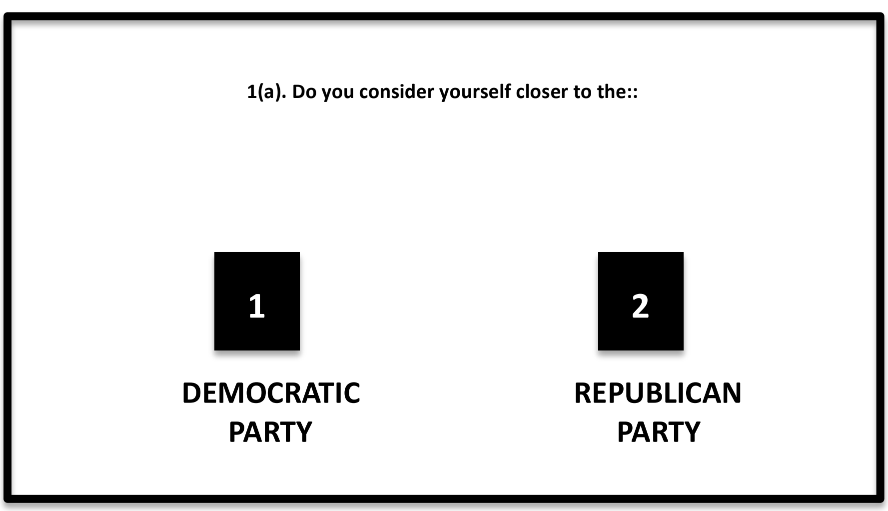
You friendly boatmen and mechanics! You roughs!

You twain! And all processions moving along the streets!

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







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).

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

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

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 $\mathrm{X}^2(k)$ 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 $S_d < \mathrm{X}^2_{.95}(k)$ are used to calculate a series of goodness of fit statistics where values of θ_d where $S_d >= \mathrm{X}^2_{.95}(k)$ 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 over all and for each type and the 1% confidence level.

We repeat the exersize 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 <u>exactly</u> 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