

CSS 205/POLI 271 Final Project

Amy Tan

March 15, 2025

1 The Original Paper

Findor et al. (2022) explored methods to increase support for out-group redistribution policies. More specifically, they aimed to test their hypothesis that distributive justice principles—justifications for who should get what and why—can increase support. In total, they ran 3 studies; however, the scope of this project will center around Study 3: Slovak Preferences over Distributive Fairness.

In Study 3, the authors recruited 1,009 Slovak citizens for an online survey experiment, using quotas on gender, age, region, size of municipality, and education to ensure the sample was representative of the national population. They then used a between-subjects design with a control group and three treatment conditions, each corresponding to a distributive principle (Equality, Reciprocity, Need). Given this information, there aren't any immediate concerns regarding the independence of participant responses - participant responses don't appear to be related to each other.

The dependent variable in Study 3 was personal welfare policy support, measured on an ordered categorical scale (1: Completely Disagree, 2: Rather Disagree, 3: Rather Agree, 4: Completely Agree). Though there was a 5th and 6th option (Don't Know and Refusal, respectively), the authors cleaned out these values, restricting data used to 1-4 with a total of $N = 1009$ observations. Figure 1 contains a bar graph visualization of the dependent variable. Interestingly, the distribution is uni-modal. The mode of 3 indicates that it was most common for participants to rather agree with welfare policies. There seems to be a relatively equal number of participants who completely disagree and agree, and the least common for participants to rather disagree with welfare policies.

The authors ultimately found that the Reciprocity condition elicited the most positive personal support. They did not make any mention of missing data in their paper or available code in relation to Study 3. Running additional checks for data missingness in the provided data set did not yield any missing values; I verified missingness by checking for NA values and confirmed that no cases were removed or missing values imputed.

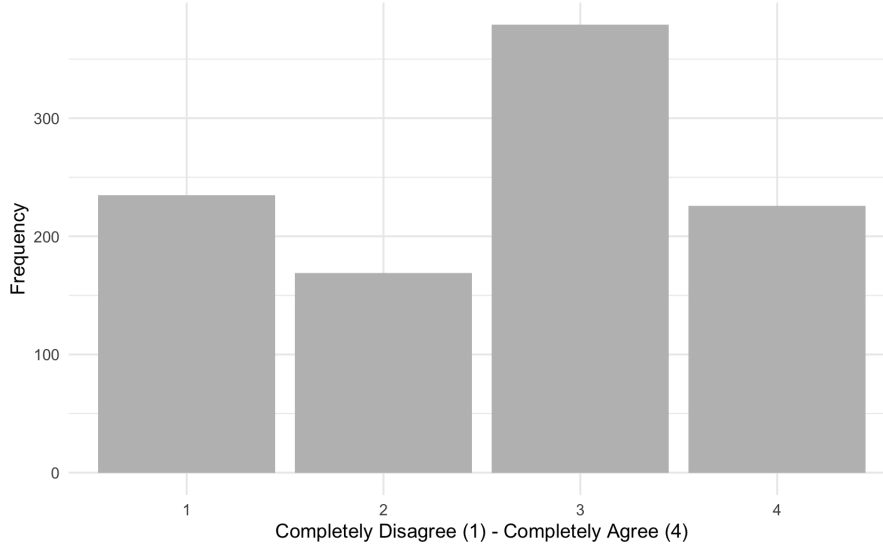


Figure 1: Distribution of Personal Welfare Policy Support

1.1 The Original Model (Model 1)

The model replicated is the ordered logit shown in the top half of Table 3 in the original paper. Participants' experimental condition (Control, Equality, Reciprocity, Need) was used to predict their personal agreement with the proposed social housing policy (measured on the 1–4 scale). The authors ran the model twice: once using the Control group as the baseline and again using Reciprocity as the baseline. Their goal with these analyses was to determine whether and how different distributive justice principles influenced support for policies benefiting minority groups, focusing on whether reciprocity significantly increased personal support relative to the control condition. By using this approach, the authors aimed to achieve causal inference about the effectiveness of different justifications for welfare policies.

I was able to successfully replicate the author's original model. Table 1 presents the results of this replication, with both baselines included.

2 An Additional Model (Model 2)

The original paper only used one predictor (the experimental group) to predict personal agreement. However, their dataset had a variety of demographic factors recorded for the participants. Thus, I was curious how adding additional predictors, specifically age, sex and education, would affect the model performance. I thought that adding these variables would lead to a model that would provide a more informative scenario regarding how an average Slovakian citizen agreed with the distributive principles.

I ran an additional ordered logit, adding these new predictors into the original model. Table 2 displays the regression table of the new model.

Table 1: Original Model Replication

	<i>Dependent variable:</i>	
	DV: Personal Agreement (1-4)	
	Base = Control	Base = Reciprocity
	(1)	(2)
Control		−0.60*** (0.16)
Equality	0.22 (0.16)	−0.38** (0.16)
Reciprocity	0.60*** (0.16)	
Need	−0.16 (0.16)	−0.76*** (0.16)
Observations	1,009	1,009
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

Table 2: New Ordered Logit

	<i>Dependent variable:</i>
	DV: Personal Agreement (1-4)
	Base = Control
Reciprocity	0.58*** (0.16)
Equality	0.20 (0.16)
Need	−0.18 (0.16)
Secondary Edu (no diploma)	−0.47** (0.22)
Secondary Edu (complete)	0.19 (0.22)
University Edu	0.45* (0.24)
Female	−0.09 (0.12)
Age	0.01*** (0.004)
Observations	1,009
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

3 Comparing Models

The question of which of the two models is best now comes to issue. Table 3 shows a summary of both models and their respective performance metrics at the bottom. Model 1 is the original model reproduced from the authors' paper, while Model 2 is the new model with the additional demographic variables.

The models' in-sample performance can be compared first. Since both models were trained using the same data ($N = 1009$) and are estimated via maximum likelihood, the log likelihood, AIC, and BIC can be compared. Larger log-likelihoods indicate that a given model is a better fit to the observed data, and is more likely to have produced the given data. Model 2 has a higher log likelihood (-1311.37) relative to Model 1 (-1341.27). The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are both additional in-sample performance metrics - they differ from the log likelihood in that they penalize for model complexity. This is especially useful considering the added complexity the additional variables add to Model 2 relative to Model 1. With both the AIC and the BIC, smaller values are preferred. Model 2 has a smaller AIC (2644.74) and BIC (2698.83) relative to Model 1 (2694.53 and 2724.03 respectively). Even with penalties for complexity, Model 2 appears preferable over Model 1.

However, the models' out-of-sample performance must also be considered. Arguably, the out-of-sample performance may be important than in-sample performance for selection, since out-of-sample performance evaluates how well models capture the underlying social and data generating processes. Thus, the out-of-sample performance for both models was conducted through cross-validation. The total dataset was randomly split (setting a random seed), with 70% used for model training and the remaining 30% used for model testing. From there, both models' accuracy on the testing data can be computed and compared. Model 2 only has a slightly higher accuracy (0.39) compared to Model 1 (0.38).

Taking both the in-sample and out-of-sample performance into account, I argue that Model 2 is preferable. Not only does it have better in-sample metrics as previously discussed, but it also has a slightly higher accuracy. While this difference is small, I argue that even if the accuracies were equivalent, Model 2 is preferable for the purpose of providing more informative and insightful scenarios as to who would be more or less likely to votes in a particular manner.

4 Interpreting Model 2

Moving forward with Model 2, we can interpret and utilize it to see how sex relates to personal support for reciprocity based welfare policy. More specifically, since reciprocity appears to be the most effective distributive justice principle, the quantity of interest for this project is the change in the

Table 3: Comparing Ordered Logits

	<i>Dependent variable:</i>	
	DV: Personal Agreement (1-4)	
	(1)	(2)
Equality	0.22 (0.16)	0.20 (0.16)
Reciprocity	0.60*** (0.16)	0.58*** (0.16)
Need	-0.16 (0.16)	-0.18 (0.16)
Secondary Edu (no diploma)		-0.47** (0.22)
Secondary Edu (complete)		0.19 (0.22)
University Edu		0.45* (0.24)
Female		-0.09 (0.12)
Age		0.01*** (0.004)
In-Sample Performance		
Log Likelihood	-1341.27	-1311.37
AIC	2694.53	2644.74
BIC	2724.03	2698.83
Out-of-sample Performance		
Accuracy	0.38	0.39
Observations	1,009	1,009
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

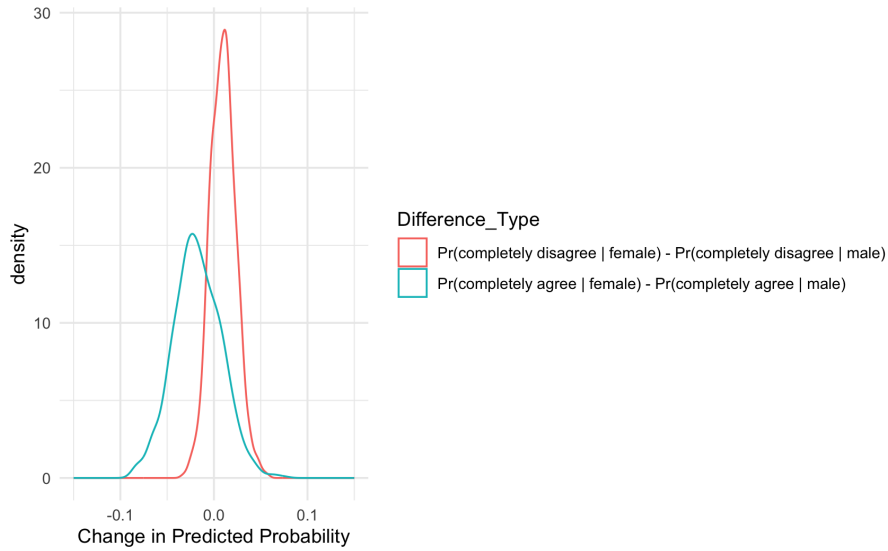


Figure 2: Change in Predicted Probability of Support of Reciprocity Based Welfare Policy based on Sex

predicted probability of completely agreeing and disagreeing with personal welfare support between males and females, holding age and education at their central tendencies (completed secondary education, 45 years of age). Figure 2 provides insight into this particular scenario.

Beginning with the change in predicted probability that an individual completely disagrees with reciprocity based welfare policy given their sex (shown in red), we see that females are slightly more likely to completely disagree than males - there is an increase in the predicted probability. While this change/difference in probability is quite small, the height of the sampling distribution provides additional confidence in this increase. But, it is worth noting that the spread of the distribution, though relatively small, contains and captures 0. Thus, there is the possibility that there is real no change or difference the probability of completely disagreeing with reciprocity based welfare policy based on sex.

Additionally, we can examine the change in the predicted probability that an individual completely agrees with reciprocity based welfare policy given their sex (shown in blue). Based on the sampling distribution mode, we see that females are slightly less likely to completely agree compared to males - there is a decrease in the predicted probability. However, in comparison to the sampling distribution of previous scenario, it is worth noting that this one has a lower density at the mode. This, coupled with the larger relative spread (that also contains 0), increases uncertainty of the conclusions that can be drawn.

Given the overall uncertainty described for both completely agreeing and completely disagreeing, it is possible that the effect of sex for both may be minimal in practice.

5 Conclusion

Overall, the authors used an ordered logit with the distributive principles as covariates in order to infer whether some principles were more effective in garnering personal support. Though there arguably might have been a need for a neutral option encoded into the Likert scale used, the use of an ordered logit was appropriate given the data the authors were working with. I would argue that rerunning the model, using Reciprocity as the baseline after already using Control as the baseline was redundant, but perhaps the authors thought the new coefficients would better show how reciprocity was more effective relative to all other conditions. I am quite confident in the authors' conclusions in Study 3 following this replication success.

My addition to this work sought to perform another ordered logit, using additional demographic variables as covariates. Using this model, selected based on in and out-of-sample performance, to gain deeper insight into the level of support the average Slovak citizen gives to reciprocity based welfare policy, we see observe (with a level of uncertainty) that females are less likely to completely agree compared to males and more likely to completely disagree compared to males.

Additional work can continue to build and select various models, as the provided datasets not only have a variety of variables unused, but other models such as multinomial logits can be experimented with as well.

5.1 Appendix

I used ChatGPT (<https://chatgpt.com/share/67af6953-53d0-8008-a46d-3be0d43b06fc>) to help break down the original papers' data cleaning process back when I initially thought I would replicate Study 2. It was quite helpful and helped me understand how they cleaned the filtered the raw data.

5.2 References

Findor, Andrej, Matej Hruška, Roman Hlatky, Tomáš Hrustič, and Zuzana Boselova. 2022. "Equality, Reciprocity, or Need? Bolstering Welfare Policy Support for Marginalized Groups with Distributive Fairness." *American Political Science Review* 117(3): 805–21. doi: 10.1017/S0003055422001046.