PRE-ANALYSIS PLAN
Costly implementation of third party preferences with non-paternalistic motivation.

Xianwen Chen[*], Øivind Schøyen[†‡]

May 24, 2017

Abstract

In an experimental setting, we plan to test whether spectators have a willingness to pay to have an income distribution adhering to their fairness views implemented upon a pair of worker. Before spectators decide whether to implement their fairness views, we elicit participants perceived prevalence of their own fairness views to see whether this is correlated with their willingness to implement. In one treatment, spectators are informed about the empirical prevalence of their fairness views among the workers, before their willingness to pay is elicited. We argue that existence of a statistically significant treatment effect would be indicative of non-paternalistic motivation of implementing fairness views.

[*]Department of Landscape Ecology, Norwegian Institute for Nature Research. Email: xianwen.chen@nina.no.

[†]The Choice Lab, Department of Economics, Norwegian School of Economics. Email: oivind.schoyen@nhh.no.

[‡]Corresponding author.

1 Introduction

A number of real world situations involve costly implementation where the incentives are dispersed to such an extent that they entail a relatively low expected benefit to the the decision maker, examples include parenting or voting (Edlin et al., 2007). This suggests that the behavior is closely linked to a willingness to exert effort to implement third-party preferences. We define third-party preference as a social preference where the decision maker does not have any private extrinsic incentive over the outcomes, but prefers other persons to have certain outcomes (Harsanyi, 1962). This type of willingness to dedicate one's resources to impose a third-party preference upon others could be rationalized by a belief that others share the same preference. For a non-paternalistic individual, willingness to impose a third-party preference upon others is limited to the extent to which they believe others share this preference. We define being purely paternalistic as preferring to impose a preference upon others regardless of others' preferences. Assuming rational and non-paternalistic decision makers, the perceived prevalence of a third party preference in an affected group should then predict willingness to implement this third-party preference.

The probability that a random person in an affected group shares a preference will trivially be given by the prevalence of the preference in the group. People's estimates of the prevalence of their own traits have consistently been found to be biased upward, which is a property commonly referred to as projection bias (Ross et al., 1977). Following the study by Ross et al. (1977), projection bias has been documented in different studies (Alicke et al., 2005; Blanco et al., 2014; Rubinstein and Salant, 2016). Due to the relevance of preference prevalence for non-paternalistic decision makers, projection bias could play a crucial role in understanding what drives people's willingness to impose third-party preferences. This is especially true of implementing one of mutually excluding alternatives adhering to contested views, examples include civil society, elections, and revolutions. To the best of our knowledge, the link between prevalence estimates and willingness of implementation has not been investigated in the literature.

We plan to experimentally test how a spectator's willingness to pay to impose her third-party preference upon workers is related to her projected prevalence of her preference in the affected group an experimental setting.[1]

We study this effect using third-party redistribution preferences. First, spectators are elicited for what they perceive to be a fair way to distribute income in a hypothetical scenario where a pair of workers have completed a task where earnings is determined by both luck and effort. The alternatives are either full redistribution or no redistribution. Spectators are then elicited for their beliefs of how many of the workers share their view. The spectators are then given the choice to pay a fee to implement the option they have reported to be the fair option upon a pair of workers that have completed a task identical to the one in the hypothetical scenario. If they choose not to pay, the other redistribution option will be implemented by default.

We seek to study the the following questions. First, we test whether people are willing to pay to implement their redistribution preferences. Second, we study whether the willingness to implement one's redistribution preference upon others is correlated

[1]We refer to experiment participant randomly assigned the role of spectator as spectators and experimental participants assigned the role of worker as workers.

with the decision maker's projected prevalence of her third-party preference. Finally, to study the extent to which third-party preferences are paternalistic, i.e independent of prevalence, we introduce a treatment where the spectators are informed about the true prevalence of their preferences before choosing whether to implement. Assuming spectators with purely paternalistic preferences, there should be no correlation between revealing the true prevalence of their fairness preference and willingness to implement the third party outcome, and consequently no treatment effect.

Beyond our main research questions, getting a better understanding of whether there is a willingness to pay to implement third-party preference is of methodological value as third-party preferences are increasingly being used in the experimental literature, see for instance Cappelen et al. (2013) and Andreoni et al. (2016).

The pre-analysis plan proceeds as follows. First we describe the planned experiment, we then describe how data analysis will be conducted before the paper concludes. The experimental instructions are presented in the appendix.

2 Experimental design

The experiment will be conducted at M:Turk. It will contain two types of roles assigned with separate tasks: workers and spectators. The spectator sessions will be the interest of the study. Spectators are elicited for beliefs and preferences. They will make a choice of whether to pay to implement their third-party preference upon a pair of workers. Workers complete a work task and are elicited for preferences.

First, a work task will be performed where a worker will be paid a show-up fee of USD 1.00 and a bonus for correctly identified words. Second, the two spectator sessions, one baseline session and one treatment sessions will be held simultaneously. Participants are randomly allocated to sessions. A spectator is paid a show-up fee of USD 1.11, which can be deducted by USD 0.02 if they choose to implement their redistribution preference.

First, there will be a worker session. Before completing the work task, the workers are elicited for their third-party fairness preferences for redistribution. Workers are then paired up and complete a task of identifying a word and letter combination from a list. Each worker draws a random price between USD 0.01 and 0.10, for each correctly identified letter. A worker will earn this price times the number of words that were correctly identified within the given time frame of one minute. Each pair of workers works for four separate rounds. In every round, a pair of workers is paired with a different spectator and paid according to the spectator's choice.

There will then be two spectator sessions: a control-group spectator session and a treatment-group spectator session. In both sessions, spectators are elicited for (1) their third-party redistribution preferences and (2) their prevalence estimates of the workers' third-party redistribution preferences.

In the control-group spectator session, the spectators are given a choice on whether to pay USD 0.02 to implement their preferred redistribution upon a pair of workers. The amount is deliberately chosen to be low, to permit identification of very "weak" third-party preferences. If they do not pay USD 0.02 to implement their preferred distribution, their non-preferred distribution will be implemented.

In the treatment-group spectator session, the spectators will be informed about the

true prevalence of their third-party redistribution preferences among the workers, before they make the choice of whether to pay USD 0.02 to implement it. This is done to investigate whether knowing the true prevalence will affect the probability of a spectator being willing to pay to implement his third-party preference.

2.1 Overview of the experimental design by session

Each spectator makes a decision for one round of work by one pair of workers. A worker works for four separate rounds and is paired with a different worker and different spectator each round. Participants are randomly allocated to be workers or to one of the two spectator treatments.

2.1.1 Sequence for workers

- Stage 1: The workers are elicited for their third-party redistribution preference for redistribution or no redistribution in a hypothetical scenario.
- Stage 2: Workers are then paired in pairs of two. Workers perform a work task consisting of identifying what number is next to a letter on a list of letter number combinations. Each worker draws a random price and earns the number of correctly identified letters multiplied by his or her drawn price.
- Stage 3: Workers are paid according to the choice of one spectator.

For each pair of workers, Stage 2-3 repeat for four rounds. In each round, the workers will be paired with a new partner and the pair will be assigned with a new unique spectator. Each spectator will decide whether to implement her preferred redistribution preference upon one pair of workers for one round.

2.1.2 Sequence for control-group spectators

- Stage 1: Spectators are elicited for their third-party redistribution preferences.
- Stage 2: Spectators are elicited for their beliefs of how many of the workers share their third-party preferences, α_i.
- Stage 3: Spectators choose to whether to pay USD 0.02 to implement their preferred redistribution option.

2.1.3 Sequence for treatment-group spectators

- Stage 1: Spectators are elicited for their third-party redistribution preferences.
- Stage 2: Spectators are elicited for their beliefs of how many of the workers shares their third-party preferences, α_i.
- Stage 3: Spectators are informed about the empirical prevalence of their redistribution preference among the workers, α_W.
- Stage 4: Spectators choose to whether to pay USD 0.02 to implement their preferred redistribution option.

2.2 Practical organization

The power analysis is done to show under what assumptions, a given sample size will give statistically significant estimate of the treatment effect. Since there is, to our knowledge, no previous experiment willingness to pay to implement redistribution preference, we do not have previous data to build on. Moreover, we do not make assumptions on standard deviations, because for power analyses of a two-sample proportions test, only proportions are needed.

The power analysis has been conducted in Stata 13.1. Our study is mainly concerned with the treatment difference between the proportions of spectators that choose to pay to impose their redistribution preference. There are two groups of spectators, from the control-group and from the treatment-group. Therefore, a power analysis of a two-sample proportions test is conducted.

Let n_{sc} and n_{st} be the numbers of spectators in the control group and in the treatment group, respectively, that we include in the experiment. We choose to have equal number of participants in the control group and the treatment group, such that $n_{sc} = n_{st}$. Assume that 90% of the spectators in the control group are willing to pay USD 0.02 to implement their redistribution preferences. Assume that 81% of the spectators in the treatment group are willing to pay USD 0.02 to implement their redistribution preferences, after learning the true prevalence of their third-party redistribution preference among the workers.

For alpha of 5% and beta of 10%, the power analysis for two-sample proportions test suggest that we need $n_{sc} = n_{st} = 320$.[2] Therefore, we will need 640 spectators. Because every spectator watches a pair of workers for each one of the eight rounds, we will need 320 workers.

Our budget allows for a slightly bigger sample size, we therefore recruit a sample size of 1008 participants, among which 336 are workers, 336 are control group spectators, and 336 are treatment group spectators.

3 Data analysis

There are three main hypothesis in the experiment:

1. Some, but not all, spectators will be willing to pay USD 0.02 to implement their preferred distribution.

2. There will be a positive association between a spectators prevalence estimates of their third-party redistribution preference α_i and their willingness to pay to implement her third-party redistribution preference.

3. Relative to the treatment effect of informing true prevalence of third-party redistribution preferences will lead to fewer spectators being willing to implement their redistribution preferences in the treatment session.

Assuming a negative treatment effect is found, it will be supportive evidence that individuals take others' fairness views into account when deciding whether to assert effort to implement their own fairness views.

[2]Stata code: *power twoproportions 0.90 0.81, alpha (0.05) power(0.9).*

3.1 Hypotheses testing

Subscripts $_{NR,R}$ denote participants with third-party preferences for the no-redistribution option and the redistribution option, respectively. Superscripts ^{BT} or ^{IT} refers to the baseline treatment (control group) and the information treatment (treatment group), respectively. The indicator variable P indicates whether a subject had a willingness to pay to implement their third-party fairness preference. Any overhead line of a variable represents the average value of this variable.

3.1.1 Hypothesis 1: Willingness to implement

The first hypothesis is that there will be a non-zero portion of spectators in both treatment and control group that are willing to pay USD 0.02 to implement their third-party redistribution preference.

The null hypothesis 1 is:

$$H_0^1 : \#[P = 1] = 0. \tag{3.1}$$

The alternative hypothesis 1 is:

$$H_1^1 : \#[P = 1] > 0. \tag{3.2}$$

3.1.2 Hypothesis 2: Correlation between α_i and implementation

The second hypothesis is that spectators, who are willing to pay USD 0.02 to implement their third-party redistribution preference, have on average a higher projected prevalence of their third-party redistribution preferences than the spectators that are not willing to pay to implement. That is, the average spectator α_i, $\overline{\alpha}$ will be higher for spectators willing to pay USD 0.02 than spectators not willing to USD 0.02. This hypothesis is tested on the entire sample of spectators, and for the sub-samples of the control and treatment group.

Entire sample of spectators For entire sample of spectators, from both the control group and the treatment group, the null hypothesis $2a$ is:

$$H_0^{2a} : [\overline{\alpha}|P = 1] \leq [\overline{\alpha}|P = 0] \tag{3.3}$$

The alternative hypothesis $2a$ is:

$$H_1^{2a} : [\overline{\alpha}|P = 1] > [\overline{\alpha}|P = 0] \tag{3.4}$$

We test this using OLS.

Control group spectators For the control group spectators, the null hypothesis $2b$ is:

$$H_0^{2b} : [\overline{\alpha}^{BT}|P = 1] \leq [\overline{\alpha}^{BT}|P = 0] \tag{3.5}$$

The alternative hypothesis $2b$ is:

$$H_1^{2b} : [\overline{\alpha}^{BT}|P = 1] > [\overline{\alpha}^{BT}|P = 0] \tag{3.6}$$

We test this using OLS.

Treatment group spectators For the treatment group spectators, the null hypothesis $2c$ is:

$$H_0^{2c} : [\overline{\alpha}^{IT}|P = 1] \leq [\overline{\alpha}^{IT}|P = 0] \tag{3.7}$$

The alternative hypothesis $2c$ is:

$$H_1^{2c} : [\overline{\alpha}^{IT}|P = 1] > [\overline{\alpha}^{IT}|P = 0] \tag{3.8}$$

We test this using OLS.

3.1.3 First-order stochastic dominance

We will check for difference in the conditional distribution of α_i; projected prevalence of distribution preferences. Specifically, we will check whether the cumulative distribution function $F(\alpha|P = 1)$ first order stochastically dominates (FOSD) $F(\alpha|P = 0)$ for the entire sample of spectators. Similarly, we will check if this holds for the sub samples of the treatment groups, i.e. α^{BT} and α^{IT}, using a Kolmogorov-Smirnov test. The null hypothesis will be no difference in distribution while the alternative hypothesis will be FOSD.

3.2 Hypothesis 3: Treatment Effect

The hypothesis is that participant on average will over estimate the prevalence of their own fairness view. The treatment consists of informing spectators about the empirical average of their fairness view amongst the workers. As we have argued in the introduction, which will be further elaborated in a brief theory section in the final paper, the utility of implementing for participants with non-paternalistic motivations sinks with lower prevalence. Hence, we expect the probability of a participant implementing will be lower in information treatment for participants for participants that have reported higher α_i estimates than the empirical α amongst the workers.
In some of the tests, we also investigate the consequences of having too low α_i estimates.

For all tests in this subsection, we first use all spectators, of both control group and treatment group. Then, we use all spectators who favored redistribution, of both control group and treatment group. Then, we use all spectators who favored non-redistribution, of both control group and treatment group.

Introducing a binary variable D_{IT} for the information treatment, we can see if the probability of paying USD 0.02 cents is lower for the group subject to the information treatment:

$$[P_i = 1] = \beta_0 + \beta_1 D_{IT} + \epsilon_i \tag{3.9}$$

We plan to run the regression for the subsample of participants with third party preferences for redistributing and for not redistributing. The hypothesis testing will be the same, for β_3 and for the two subsample regressions:

$$H_0^{5a} : \beta_1 \leq 0 \ , \tag{3.10}$$

while the alternative hypothesis is that:

$$H_1^{5a} : \beta_1 > 0 \ , \tag{3.11}$$

Because we expect that the treatment is stronger for participants with higher α estimates, we can also test the interaction effect. Therefore, we plan to run the following OLS regression:

$$[P_i = 1] = \beta_0 + \beta_1 \alpha_i + \beta_2 D_{IT} + \beta_3 [D_{IT} \times [\alpha_i - \alpha_W]] + \epsilon_i \tag{3.12}$$

We plan to run the regression for the subsample of participants with third party preferences for redistributing and for not redistributing. The hypothesis testing will be the same, for β_3 and for the two subsample regressions:

$$H_0^{5b} : \beta_3 < 0 \ , \tag{3.13}$$

while the alternative hypothesis is that:

$$H_1^{5b} : \beta_3 \leq 0 \ , \tag{3.14}$$

We will explore differences in the probability distribution of having a willingness to implement contingent on the α_i values for the control and treatment group. We expect that for the majority of the sample the treatment group will have systematically lower willingness to implement and that some minor portion of the sample will have a downward bias in their α_i value such that the treatment effect will increase their willingness to implement.

We will also explore whether we find evidence on an non-linear treatment effect on willingness to implement and having a particular high or low estimate of own fairness view.

4 Conclusion

The planned experiment will test for whether prevalence estimates predict willingness to pay to implement spectators preferred third-party preference. Further, we will test whether correcting for inflated prevalence estimates will reduce willingness to pay to implement a third-party preference. Finding a treatment effect will be an indication that people do not have paternalistic preferences. A lower willingness to pay to implement a preferred distribution for spectators with inflated prevalence estimates would also indicate that projection bias plays an important role in explaining voluntary participation in activities aimed at imposing contested states of the world upon third parties.

A Appendix A: Experimental instructions

All text in italics is left out of the experiment and only included for a reader overview. Screens indicate when the program will change text.

A.1 Worker instructions

A.1.1 *Screen 1: Introduction*

Thank you for your participation in this experiment. Please read all instructions carefully.

A.1.2 *Screen 2: Hypothetical Scenario*

Assume two workers have been completing an identical task. The task is identifying the number on a list that is next to a given letter. The worker gets paid per correctly identified word. Each worker gets paid a separate randomly drawn price. The price can be any whole number from 1 to 10 cents.

Elicit hypothetical preferences Which of the payment options do you find to be the most fair option?

1. No-redistribution: Each worker is paid separately for their work. In other words the workers get paid for the number of words they identified time the price they are randomly assigned.

2. Redistribution: The total earnings of the two of workers are divided equally among the workers. In other words each workers gets paid the sum of the payments of the two workers divided by two.

A.1.3 *Screen 3: Work task*

You are now going to perform a letter-number decoding task equal to that described in the scenario. A sequence of letters with corresponding numbers will be displayed on the screen. You should write the number corresponding to the given letter in the box below the sequence.

An example of the task is provided below *(Figure 1)*.

A new sequence will be displayed directly below the first. You will not know whether your answers are correct until the end of the task. The task will last for 60 seconds.

Remaining time will be displayed at the top of the page.

After completing the task a price between 1 and 10 cents per word will be drawn at random. Another experimental participant will choose whether your earnings will be redistributed with another worker, or whether you will receive payment equal to your earnings.

A.1.4 *Screen 4-9*

Five letter identifier list will appear for the experimental subjects.

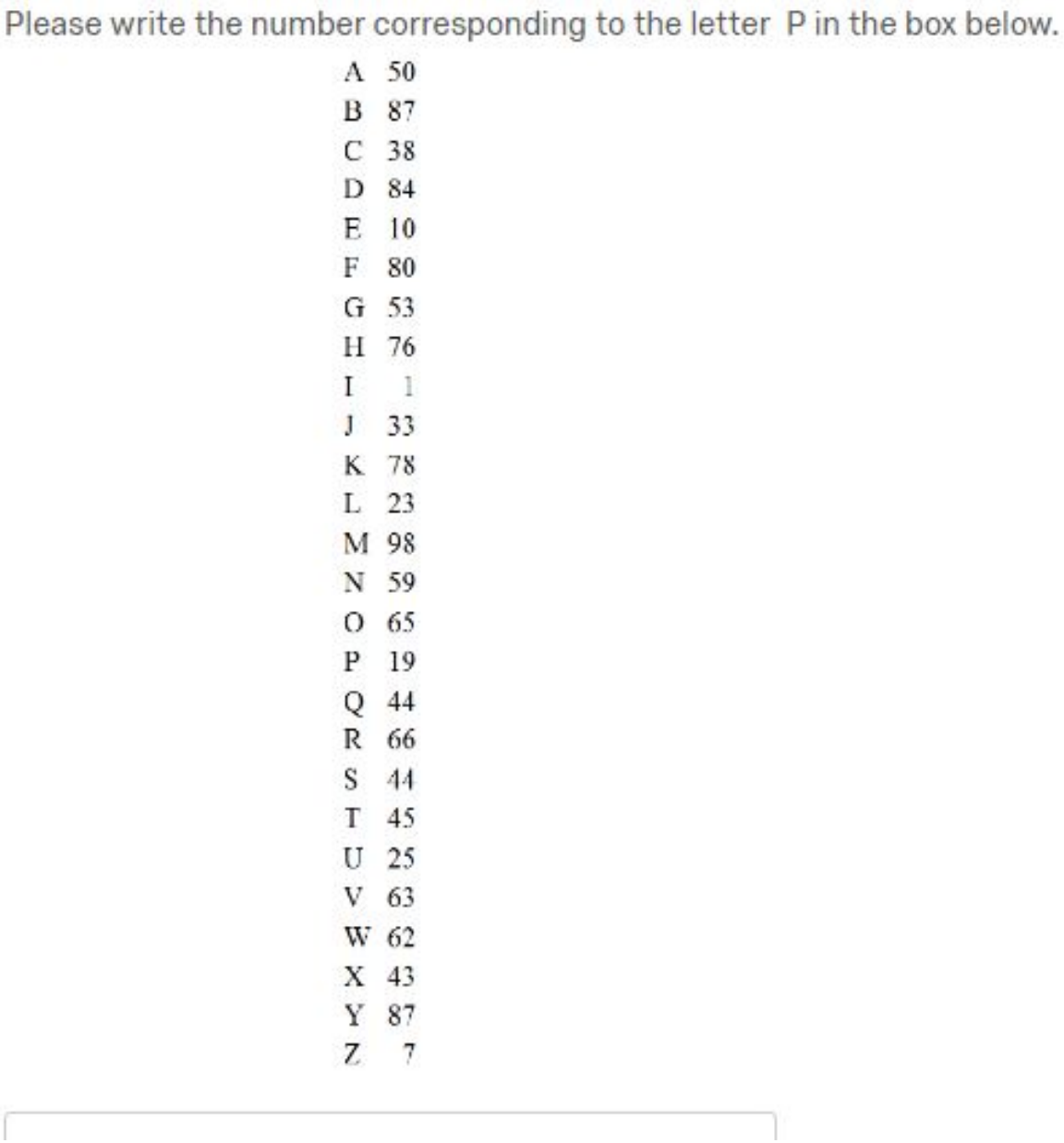

Figure 1: Example of letter identifier list.

A.1.5 *Screen 10: Information of payment*

Thank you for participating. You have identified X numbers at the drawn price of Y your earnings is Z. You will now participate in a second identical round of work.

This repeats for eight rounds of work

A.1.6 *Screen 56: Information of payment*

Thank you for participating. You have identified X numbers at the drawn price of Y your earnings is Z. You will be informed at a later stage about what your final payment will be.

A.1.7 *Screen 57: End*

Thank you for your participation!

A.2 Spectator session: control-group

A.2.1 *Screen 1: Introduction*

Thank you for your participation in this experiment. Please read all instructions carefully.

A.2.2 *Screen 2: Hypothetical Scenario*

Assume two workers have been completing an identical task. The task is identifying the number on a list that is next to a given letter. The worker gets paid per correctly identified word and each worker gets paid a separate randomly drawn price. The price can be any whole number from 1 to 10 cents.

Elicit hypothetical preferences Which of the payment options do you find to be the most fair option?

1. No-redistribution: Each worker is paid their separate earnings. In other words the workers get paid for the number of words they identified time the price they are randomly assigned.

2. Redistribution: The total earnings of the pair of workers are divided equally among the two workers. In other words each workers gets paid the sum of the payments of the two workers in the pair divided by two.

A.2.3 *Screen 3: Real Scenario*

There has been performed an experiment like the one described on the M:Turk platform with participants like yourself. 335 worker participants completed the task as described and drew a random price between 1 and 10 cents. Before completing the work task the workers were asked which of the payment options, redistribution or no-redistribution they found to be the fair option.

Elicit α: for participants with third party preferences for redistribution Among the 100 worker participants how many do you think find the redistribution option fair?

Elicit α: for participants with third party preferences for no-redistribution Among the 100 worker participants how many do you think find the no-redistribution option fair?

A.2.4 *Screen 4A: Elicit real 2 cent preferences for participants with third party preferences for redistribution*

You are to decide how payment should be done for one pair of worker participants. You can pay 2 cents of your 1.11 dollar participation earning to implement the distribution you previously stated you found to be the fair option; the redistribution option.

I choose to:

1. Pay 2 cents to implement the redistribution option.

2. Not pay 2 cents. The no-redistribution option will then be implemented.

A.2.5 *Screen 4B: Elicit real 2 cent preferences for participants with third party preferences for no-redistribution*

You are to decide how payment should be done for one pair of worker participants. You can pay 2 cents of your 1.11 dollar participation earning to implement the distribution you previously stated you found to be the fair option; the no-redistribution option.

Would you like to pay 2 cents to implement the no-redistribution option?

I choose to:

1. Pay 2 cents to implement the no-redistribution option.

2. Not pay 2 cents. The redistribution option will then be implemented.

A.2.6 *Screen 5: End*

Thank you for your participation!

A.3 Spectator session: treatment-group

For the treatment-group spectator session, everything is the same as in the control-group spectator session. The only difference is in the following.

Treatment text before eliciting willingness to implement preference at screen 4 reads: Treatment Screen: Revelation of empirical α Among the 100 worker participants XX preferred the redistribution option.

References

Alicke, M. D., Dunning, D. A., and Krueger, J. (2005). *The self in social judgment*. Psychology Press.

Andreoni, J., Aydin, D., Barton, B., Bernheim, B. D., and Naecker, J. (2016). When fair isn't fair: Sophisticated time inconsistency in social preferences. *Working paper*.

Blanco, M., Engelmann, D., Koch, A. K., and Normann, H.-T. (2014). Preferences and beliefs in a sequential social dilemma: a within-subjects analysis. *Games and Economic Behavior*, 87:122–135.

Cappelen, A. W., Konow, J., Sørensen, E. Ø., and Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, 103(4):1398–1413.

Edlin, A., Gelman, A., and Kaplan, N. (2007). Voting as a rational choice why and how people vote to improve the well-being of others. *Rationality and Society*, 19(3):293–314.

Harsanyi, J. C. (1962). Bargaining in ignorance of the opponent's utility function. *Journal of Conflict Resolution*, pages 29–38.

Ross, L., Greene, D., and House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3):279–301.

Rubinstein, A. and Salant, Y. (2016). " Isn't everyone like me?": On the presence of self-similarity in strategic interactions. *Judgment and Decision Making*, 11(2):168.