SIEPR Discussion Paper No. 15-021

The Lives of Others: Predicting Donations with Non-Choice Responses

By

Jeffrey Naecker

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Lives of Others: Predicting Donations with Non-Choice Responses

Jeffrey Naecker*

April 15, 2015

Abstract

There is significant variation in the percentage of adults registered as organ donors across the United States. Some of this variation may be due to characteristics of the sign-up process, in particular the form that is used when state residents renew or apply for their driver's licenses. However, it is difficult to model and predict the success of the different forms with typical methods, due to the exceptionally large feature space and the limited data. To surmount this problem, I apply a methodology that uses data on subjective non-choice reactions to predict choices. I find that active (ie yes-no) framing of the designation question decreases designation rates by 2-3 percentage points relative to an opt-in framing. Additionally, I show that this methodology can predict behavior in an experimental setting involving social motives where we have good structural benchmarks. More generally, this methodology can be used to perform policy pseudo-experiments where field experiments would prove prohibitively expensive or difficult.

Keywords: organ donation, social preferences, lab experiment

JEL: C91,D12,H31,Q51

*Department of Economics, Stanford University; jnaecker@stanford.edu. I thank Doug Bernheim, Al Roth, Muriel Niederle, Charles Sprenger, and Judd Kessler for their irreplaceable advice. I thank Judd Kessler and Linda Yao for providing the DMV forms used in my analysis. I also thank attendees at the Stanford Behavioral and Experimental workshop, the 2014 Stanford Institute for Theoretical Economics, and the 2014 Economic Science Association meetings.

1 Introduction

Demand for organ donation is high in the United States: There are currently over 120,000 people awaiting transplants, and over 10,000 new candidates are added each year. Despite this large and growing need, there is a limited supply of organs available; as a result, in 2013 over 6,000 people died waiting for transplants.[1] There are two sources of donor organs: live donors, who choose to give an organ (usually a kidney) to a person in need; and deceased donors, who have agreed to donate multiple organs in the event of their death.[2] There have been several efforts to improve the supply of live donor organs[3] , but Kessler and Roth (2012) point out that a more effective policy margin may be to increase the number of potential deceased donors. The *organ donor designation share* is defined as the percentage of the adult population in the state that is registered to be an organ donor in the event of their death. There is much variation in designation shares across the United States, ranging from 19% in Texas to 80% in Alaska, with a national average of 47% as of 2012.[4] These rates are on the whole perhaps surprisingly low, given that designating oneself as a donor usually only requires one to visit a website or to check a box when applying for a driver's license. This discrepancy in shares raises the several questions: What is the source of this interstate variation in donor designation? Given the clear need for more transplantable organs, how can these rates be increased?

The differences in designation rate may be at least partly driven by differences in how the organ donation decision is elicited. For example, consider the forms used in Connecticut and Louisiana when a resident applies for a driver's license (Figure 1). Both forms use what can be called an *active choice* frame: Action is required if the respondent desires to respond in the affirmative or in the negative. That is, the lack of a response is not the same as a "No." Compare these to the form used in New Hampshire, which uses an *opt-in* choice architecture: in this case, a lack of response is indistinguishable from a "No." The active choice framing could lead to higher donor designation rates, if it prevents people from skipping or ignoring the donation question (Thaler, 2009). However, Kessler and Roth (2014b) point to experimental evidence that mandated choice may not increase donation rates, and may even have negative effects on recovery of organs post-mortem from eligible donors who have not given consent.

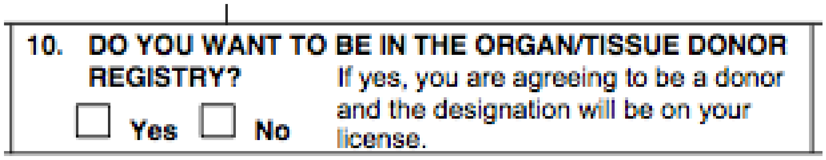
In this paper, I use a new methodology, called non-choice revealed preference, to estimate the effect of choice framing on donor designation choices. This methodology employs responses to hypothetical and subjective questions to make predictions about previously unobserved settings. When applied to the DMV forms, this methodology finds a small, negative (approximately -2%)

[1]See Organ Procurement Transplantation Network (2014) for more detailed data.

[2]Only a small percentage of those who register as deceased donors are able to donate their organs, due to the strict medical requirements for the organs to be received in good enough condition for transplantation.

[3]Some of the most successful efforts include novel market design, such as directed donation chains (Ashlagi et al., 2012).

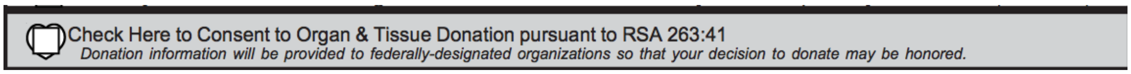
[4]See Donate Life America 2013 Report Card for more details.


10. DO YOU WANT TO BE IN THE ORGAN/TISSUE DONOR REGISTRY?
☐ Yes ☐ No
If yes, you are agreeing to be a donor and the designation will be on your license.

(a) Question on the form used in Connecticut. Note the active choice framing of the question, as well as the relatively precise and detailed wording.


ORGAN DONOR? ☐ Yes ☐ No

(b) Question on the form used in Lousiana. This question also uses active choice framing but has an extremely minimalist wording.


Check Here to Consent to Organ & Tissue Donation pursuant to RSA 263:41
Donation information will be provided to federally-designated organizations so that your decision to donate may be honored.

(c) Question on the form used in New Hampshire. This is an opt-in frame. Note also the heart symbol used to draw the user's attention to this question, and perhaps prime a more generous response.

Figure 1: Relevant snippets of forms used in Connecticut (top), Louisiana (middle), and New Hampshire (bottom) when residents apply for a new driver's license.

effect of switching from opt-in to active choice framing, a result which is in line with the best existing empirical and experimental evidence on this question. The non-choice approach, however, can be applied much more cheaply than an experiment and can give reliable estimates even when a standard econometric approach might not.

To understand why the non-choice approach might be advantageous, it is important to understand that the DMV forms differ across states in a variety of other potentially consequential ways: whether any additional information is given about the organ donation choice, whether the form asks for a monetary donation, and whether the donation question is visually highlighted. The behavioral and experimental economics literatures have shown that differences in context and information provided can have strong effects on behavior that are not predicted classical models. A recent example is found in Coffman et al. (2013), who show that changing one line in a letter to applicants accepted to Teach for America has a significant effect on their willingness to both join and stay in the program. If such framing and information architecture effects are indeed important factors in the designation decision, states could make meaningful improvements in their designation rates by simply changing the forms used at the DMV. Importantly, many such changes could presumably be implemented without the need for any new legislation.

How can we go about understanding which features of these forms have the largest effects on donor choices? The standard approach might work as follows: We could collect all available forms from across the United States and categorize them according to the objective dimensions we think are likely to matter. We could then run a regression of the designation rate on these form-specific variables, including controls for state-level variables that we think might effect designation rates. The coefficients in this regression could then be used for prediction: The coefficient corresponding to the choice framing, for example, could tell a policy-maker whether to expect an increase or decrease in designation rates in that state from a change in the choice framing.

There are several potential weaknesses with this approach, however. The first two difficulties could be said to be problems with modeling. First, it is difficult to know *a priori* which differences across state forms are the important ones. When evaluating a new form, we may simply not code a variable that turns our to be an important factor in the designation process. Is color or font choice important, for example? What about wording, placement of the question on the page, or highlighting of the organ donation question? All of these factors could make a difference for individual decisions, but including them all in a statistical model might lead to over-fitting. Second, even if we know what factors to include in our model, we may not know how to properly code them. How does one systematically describe something as subjective as how the wording of the question motivates the individual? Certainly this affects the salience or informational content of the question, but it is difficult to describe rigorously. The third potential problem with the objective data approach is one of identification. Even if a reliable model were estimated, the available data might have no variation along a key dimension, and we might be unable to estimate the effect of that variable. For example, suppose that all existing forms used only the opt-in choice framing. If we believe (from theory, laboratory experiments, or observation in related domains) that choice framing matters in this setting, how would we predict the expected effect of a switching to an opt-in frame? The necessary variation is simply not available in the data.

Such a situation is idea for application of the methods first developed in Bernheim, Bjorkegren, Naecker and Rangel (2013). That paper proposes to solve these critical issues by creating a synthetic experiment in non-choice response data, using regularization techniques to parsimoniously describe the complex mapping from non-choice response to economic activity. If identified mapping is valid, the treatment can be recovered for a proposed policy variation. This methodology may get around several potential pitfalls of the standard approach, which I will explain in the following section.

The ultimate goal of this paper is to show how this new approach might be useful for improving organ donation policy. I do this by running a "pseudo-experiment" on active and passive choice forms, through the creation of counterfactual versions of each form that differ only in their choice framing. The non-choice approach finds evidence that active choice framing has a small negative effect on the donor designation rate. Unfortunately, it is unlikely that we will observe the counterfactual forms used by the states, so the predictions cannot be directly validated. As a robustness check, I apply the methodology to an experimental setting where I can test the non-choice models out of sample more easily. I show that the non-choice approach accurately predicts behavior in previously unobserved settings, and that these predictions can be used to generate accurate estimates of theoretical parameters that are impossible to estimate with the observed choice data.

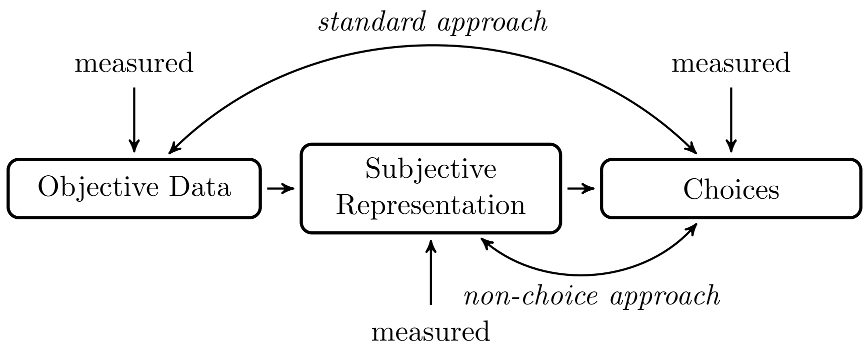


Figure 2: The model of cognition proposed in this paper. The standard economic approach (top) is to completely bypass the subjective representation. In contrast, it may be advantageous to measure proxies for the subjective representation (bottom).

2 Methodology

In this section, I explain the methodology in more detail, and review how previous research has used non-choice responses to predict choices. Consider at generalized prediction setting: We have data on choices in domain A, and we are interested in prediction choices in domain B. As discussed above, we could collect objective descriptors of the options in domains A and B, estimate a model that predicts choices from these characteristics using domain A data, and make predictions in domain B by plugging in the appropriate objective data to these models. The non-choice reactions approach has the same structure, but uses as predictors several *non-choice reactions*: unincentivized responses to questions about options or choice problems. These include questions about the subjective attributes of the options being considered, such as social image, warm glow, or moral concerns: "Which option do you think more effectively addresses a typical person's feeling good about themselves?" or "How how hard was it for you to find the organ donation question?" These non-choice questions may also include hypothetical choice questions, which can be thought of as potentially aggregating several subjective dimensions at once.

Importantly, these non-choice reactions can be collected from individuals who are not the individuals whose choices we wish to predict. All that is necessary for this to work is that the individuals providing the predictors are able to give non-choice reactions that are correlated with choices of the target group, and that this correlation structure does not change from domain A to domain B.

Why might these non-choice reactions allow us to make predictions about choices? As in Bernheim et al. (2013), I propose a two-step model of cognition that would support such a mapping from non-choice responses to choice. First, the decision-maker observes the many objective characteristics of the choice problem and translates these into a representation of the choice problem in a subjective space. Dimensions of this subjective space include visceral reactions such as which option will make her feel happy or safe, what others will think of her choice, and so on. The decision maker then makes a choice by optimizing a utility function that takes these subjective dimensions, not the original objective characteristics, as inputs. This process is summarized visually in Figure 2. The standard approach in economics is to measure choices and objective inputs, and estimate a model that relates these objects. In the non-choice response approach, one instead estimates a model that relates the subjective dimensions to outcomes.

Once the non-choice responses are collected, the researcher can estimate a model that relates these responses to choice. It is important to note that there may be a large number of potential predictors to include, for two reasons. First, because the non- choice reactions are collected from individuals but predictions are made a the choice problem level, responses must be aggregated. Many of the non-choice questions ask for responses on a Likert scale, meaning that there are many degrees of freedom in aggregating responses. For example, the research may look at the mean responses on a scale of 1 to 7, or the percentage of subjects responding with a 4 or higher. The non-choice approach considers all of these moments of the distribution of non-choice reactions as possible predictors of choice. Second, there is a large number of potentially important subjective dimensions, but only some of these dimensions may be correlated with choice. While the researcher may be able to make a judgment about which variables should be included in the model, they may include more variable than are relevant. To select the most useful set of predictors, and to avoid over-fitting on the estimation data, I perform three different types of model selection when estimating the non-choice models.

The first model selection technique is the Lasso algorithm, which solves the following minimization problem:

$$min_\beta \sum_i (y_i - \beta x_i)^2 + \lambda|\beta|$$

The first term of the loss function is the standard squared error, as in OLS. The second term penalizes for the total magnitude of coefficient used by the model. The parameter λ governs the trade-off between the two terms in the objective function. It is set by a cross-validation procedure: The data used for estimation is partitioned into a small number n of folds, and a large set of λ's is chosen for consideration.[5] For each level of λ, the model is estimated n times, each time holding out one fold. The MSE on the held out folds is then calculated. The λ for the prediction analysis can then be selected by examining the MSEs; by convention, the largest lambda with MSE within one standard deviation of the minimal MSE is usually chosen. For more details, see Zou and Hastie (2005).

The second model selection method searches for the model with the lowest Bayesian Information Criterion (BIC) on the one mover game data. This algorithm does a stepwise search of the set of possible models until it finds a local minimum in the BIC (that is, neither adding not removing a

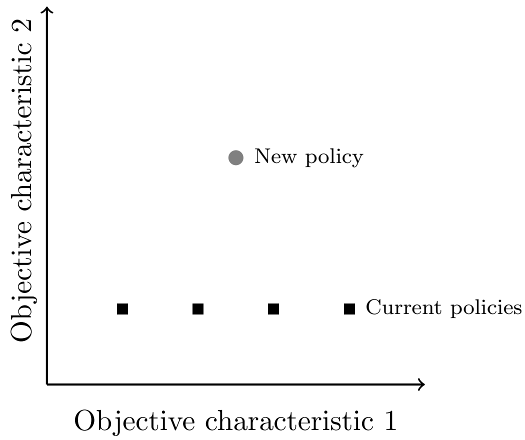
[5]The set is guaranteed to be bounded because for λ small enough, the problem reduces to the OLS problem, and for λ large enough, only the empty model (ie only a constant is included) will satisfy the minimization problem.

variable improves the BIC). The third technique is similar, except that it searches for a model with the lowest cross-validated MSPE. It is important to stress that these methods, as well as the Lasso model, do all of their model estimation selection on domain A only. Non-choice data from domain B is then applied to the resulting models to generate predictions about behavior in that domain.

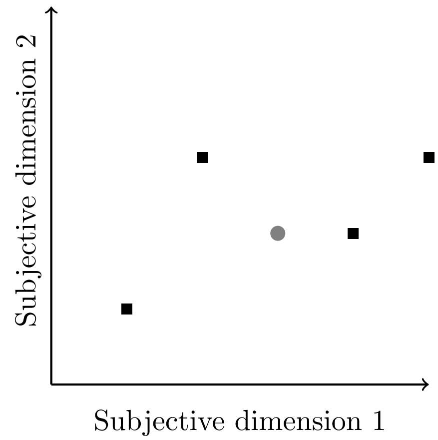
For comparison of predictive accuracy, we can calculate several statistics. Let p be a vector of predictions for J outcomes, and r be the vector of corresponding actual values. The *bias* is then defined as $\frac{1}{J}\sum_{j=1}^{J}(p_j - r_j) = \bar{p} - \bar{r}$, and is simply the average prediction error. The *mean squared prediction error* is defined as $\frac{1}{J}\sum_{j=1}^{n}(p_j - r_j)^2$. This metric is included because it is possible for a predictor to have zero *average* bias but very large absolute bias. Both the bias and the MSPE should ideally be close to zero. A less standard prediction metric that I will employ is used in Bernheim et al. (2013), called the *calibration score*. This is slope of a regression of actual on predicted values, ie β from the regression $r_j = \alpha + \beta p_j + \varepsilon_j$. A calibration score of closer to 1 is desirable: this indicates that difference in the prediction are unbiased for the actual changes in the target variable.

This approach has several advantages over the standard methodology. First, the researcher does not need to worry about capturing or coding all of the objective characteristics in her model, because the model being estimated does not use these dimensions directly as inputs. This is because the subjective dimensions are assumed to completely encode and subsume all relevant objective dimensions. This feature addresses the modeling weaknesses of the standard approach mentioned above. Second, the set of choice problems (in this case the set of DMV forms) needs to span the important subjective dimensions, but not necessarily the set of objective dimensions. For example, recall the case where all the DMV forms that the researcher has access to use only the opt- in choice framing. In this case, directly estimating the effect of choice framing would be very difficult. However, it may be that this set of available forms nonetheless spans the subjective space. That is, what matters is not the choice framing, but how that framing effects the key dimensions of the subjective representation, such as how each frame interacts with concerns about self-image, warm glow, and other motivations for pro-social behavior. Thus, through careful selection of non-choice questions, the researcher can circumvent the identification issue that may limit the standard approach. Using these non-choice responses, the researcher can still make predictions about counterfactual forms simply by collecting non-choice responses about these made-up forms.

There are several potential disadvantages with this new approach. Broadly, these concerns pertain to the stability of the non-choice model: Can we use a model that relates choice and non-choice reactions on a set of choice problems that the model has not seen? This problem has at least two specific manifestations. First, since we can't measure the subjective representation directly, we must assume that some function of non-choice reactions proxies for the subjective representation. Second, it may turn out that the non-choice questions in our data don't span the entire subjective space. We can take steps to combat these weaknesses, however. First, we can attempt to capture all



(a) Variation in the objective data in this case does not allow for identification of the efect of objective dimension 2.



(b) Variation in the subjective representation, however, does allow for identification of the effect of all key subjective dimensions. Furthermore, choice problems that appear very different in the objective space may be subjectively very similar.

Figure 3: A graphical representation of on possible failure of identification in the objective subjective representations. The black squares indicate observed data, while the gray circle indicates a choice problem or economic setting whose outcome is unobserved.

pertinent subjective dimensions by collecting responses to a large number of non-choice questions. Through careful model selection, we can select the variables which are most useful for out-of-sample prediction. And through cross-validation, we can directly test whether the non-choice models can make useful prediction in previously unseen settings.

2.1 Previous Literature on Non-Choice Responses

Non-choice responses have been used across the social sciences. In economics, the stated preference (SP) literature developed when researchers became interested in predicting choices in non-market settings where no choice data was available.[6] Hypothetical choices and stated preferences are usually biased, however, typically in the direction of a higher stated willingness to pay for goods. (See List and Gallet (2001) for a review of how hypothetical bias depends on context and elicitation method.) Two methods have been proposed in this literature to deal with this bias. First, some research has sought to ask questions that introduce less bias in the first place. One example of this approach is a cheap talk script (used by Cummings and Taylor [1999]), which encourages the respondents giving the unincentivized answer to attempt to give as close a response to their true preferences as possible. Alternatively, many researchers have sought to calibrate the hypothetical statements ex-post. This was first done by Kurz (1974), and Shogren (2006) provides a recent overview. Recent research in economics has begun to examine how well questions other than hypothetical choice or stated

[6]This literature got its name from the fact that preferences were *stated* rather than *revealed* by observation of choices.

preferences are correlated with choice. Benjamin et al. (2014), for example, compare subjective responses about well-being to actual choices of medical students applying for residency. They find that while the subjective rankings predict choice rankings well, the two data sources imply different trade-offs on key features of the residencies. This suggests that multiple non-choice data sources may be needed to fully recover underlying preference parameters.

The marketing literature has developed techniques similar to the stated preference literature. Many marketing papers use calibration techniques to examine how stated preferences predict purchase behavior. (See Chandon et al. (2005) for a recent meta-analysis of this literature.) Some papers in the marketing literature, most recently Alpízar et al. (2003), have used stated preferences to estimate structural parameters of choice functions. I perform a similar exercise in this paper, but with fitted values from non-choice models.

Research in marketing, as well as in the SP literature, has focused on predicting the responses of individuals. The non-choice approach used in this paper, on the other hand, makes predictions about aggregate behavior only. This is for two reasons: First, we imagine that the policy goal is to predict the behavior of a population under new, i.e. previously unobserved, economic conditions. Second, the subjective dimensions that are key to identification in this approach are imagined to vary on the choice-problem level. We could attempt to make predictions at the level of the individual person for each context, but making aggregate predictions of behavior under a new policy is sufficient form most applications. A further difference between this paper and the aforementioned literature is that I do not collect choices and non-choice responses from the same individuals; to do so might invite contamination of the two data sources. Again, this could be a problem in application, where the effect of a totally new economic policy is to be predicted from the responses of individuals who have never made choices in that setting, but the non-choice model is estimated using responses from individuals who have made real choices in observed settings.

The neuroscience literature has also examined the relationships between choices and non-choice responses, both in terms of brain images and more traditional non-choice data. A recent paper by Kang et al. (2011) shows that the parts of the brain activated by real choices are also activated by hypothetical ones, although the degree of activation is different. Additionally, Lebreton et al. (2009) show that brain activity while giving subjective ratings of items is correlated with hypothetical choices over those items. Smith et al. (2014) perform a predictive exercise most similar to that of this paper: they use brain imaging to predict choices of individuals over new goods, or the choices of entirely new groups of subjects. Taken together, these papers indicate that hypothetical, subjective, and choice responses all use the same valuation pathways in the brain to some degree, and that the relationship between these types of responses are at least somewhat stable across individuals and contexts. While I will not use neural data in this paper, these results are useful in that they give some credence to the existence of a subjective representation of choice problems that may be useful for prediction.

The political science literature deals very often with non-choice data in the form of opinion and voting polls. As in the stated preference literature, papers in this area fall into two different approaches. Much work has been done calibrating stated preferences from poll responses to predict voter activity (Jackman (1999), Katz and Katz (2010)). A recent paper by Rothschild and Wolfers (2011) attempts to fix the question ex ante; they find that asking voters who will win is a better prediction of the ultimate winner of an election than asking about intended voting behavior. The authors give the intuition that the expectations question gives more information: it reveals not just the voter's preferences but also their beliefs about the preferences of others. In Bernheim et al. (2013), we show that "third party hypotheticals", which ask about the likely behavior of others, can be more useful for prediction that the standard first-person hypothetical.

3 Application to Organ Donation

Recall the two choice frames commonly used on DMV forms: In the opt-in framing, a non-response is indistinguishable from a negative response. In the active choice framing, respondents must answer "Yes" or "No", and thus a non-response is not the same as a negative response. In this section, I apply the non-choice methodology to estimate the effect of choice framing on the designation rate. To judge how well the non-choice data answer this question, I will benchmark the approach with a more standard panel regression analysis, which I discuss first.

3.1 Panel Regression Approach

One standard regression-based approach to estimate the effect of choice framing is to collect the available forms from other states, and estimate a model that relates objective characteristics of these forms to the state's designation rate. As discussed earlier in the paper, however, this approach may fail if we do not capture all important objective characteristics in their model, or if there are endogeneity or selection issues.

3.1.1 Data

Donor Designation Rate The outcome variable we are interested in is the donor designation rate at the state-year level. The designation rate is defined as the percent of eligible adults applying or renewing their license or identification at the state's Department of Motor Vehicles (or similar institution) who indicate that they want to become or continue to be an organ donor. The rate for a large percentage of states is made public by Donate Life America (DLA). The rate is missing for some state-years because it was not provided by the state to DLA. I have obtained all annual rates made available from 2007 through 2012. Figure 4 plots the rate for each state over these years. The annual rates are plotted relative to the rate in 2007 (or first available year) for each state. From this figure we can see that designation rates are generally rising across the United States,

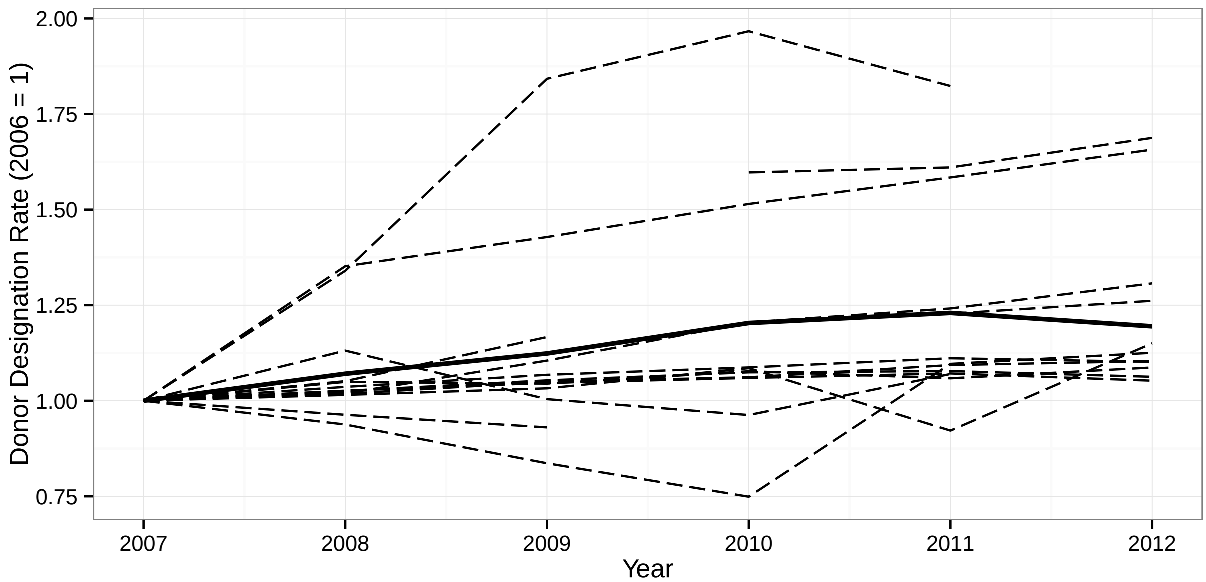


Figure 4: Growth in donor designation rate for all states individually (dashed lines) as well as the United States overall (thicker solid line). The designation rate is the fraction of license applications and renewals at the DMV that resulted in a donor designation. All rates normalized so that the rate in 2007 (or the first year of data for that state) is equal to 1. Not shown is the very large increase in Michigan, which went from a rate of less than 10 percent in 2006 to 48 percent in 2012.

though there is some heterogeneity and instability in the increase. The summary statistics for the designation rates by year are provided in Table 1, where we can see that the average designation rate in the United States increases by about 2 percentage points per year.

Objective Predictors The majority of the objective predictors are observable characteristics of the forms used as the state DMVs. I obtained 46 such forms used by 39 states.[7] I coded all the forms according to the following binary variables:

- *Active*: Whether the donation question is framed as an active choice (as opposed to an opt-in framing). 30 forms have this feature.
- *Highlight*: Whether the donation question is visually highlighted (i.e. by a different text size or color). 15 forms have this feature.
- *Money*: Whether the form also asks for a monetary donation to support organ donations. 15 forms have this feature.
- *Info*: Whether any additional information about organ donation is given on the form. 17 forms have this feature.

[7] I am deeply indebted to Judd Kessler and Linda Yao for providing copies of these forms.

Year	Mean	Median	Min	Max
2007	36.8	36.2	8.9	67.5
2008	40.8	42.9	9.4	65.0
2009	41.3	43.2	11.1	64.4
2010	42.9	44.5	12.2	76.0
2011	44.1	43.7	13.3	76.6
2012	47.4	46.0	12.7	80.3

Table 1: Summary statistics of state-level designation rates across the United States. The mean is not population- or donor-weighted. The designation rate is the fraction of license applications and renewals at the DMV that resulted in a donor designation.

Additionally, it may be the case that social norms vary by state, and this leads some some states to have higher donation rates. As a proxy for state-specific norms, I collected the average percent of household income donated in each state in 2010. Because data are not available for most of the time window of this study, this variable did not vary over time.

3.1.2 Methodology

I run the following regression:

$$DesignationRate_{it} = \beta_0+\beta_1 Active_{it}+\beta_2 Highlight_{it}+\beta_3 Money_{it}+\beta_4 Info_{it}+\beta_5 PercentGiven_i+\varepsilon_{it},$$

where i indexes states and t indexes years. I include year fixed effects, and errors are clustered at the state level. Note that I do not include state fixed effects, as most states do not change their forms over the time of the panel. As a result, any identification of the effect of the choice framing (or any other form variable) will occur across states rather than within.

3.1.3 Results

In Table 2, we can see the results of a OLS regression of the model above. In the first with the greatest coincidence of available designation rates and available DMV forms. First, we can see that the effect of the changing from opt-in to active framing is positive but not significantly different from zero at standard levels. For comparison, Kessler and Roth (2014a) estimate the effect of California switching frame from opt-in to active to be between -2.2 and -2.7 percentage points. These results are not necessarily in conflict, as my panel regression is attempting to identify the average effect of choice framing over all the states in the sample, not just California's. However, it should be clear that the identification strategy in my regression is much more tenuous than Kessler and Roth's, who use a differences-in-differences approach.

	2006-2012
(Intercept)	0.360 (0.062)***
Active	0.036 (0.028)
Highlight	−0.008 (0.030)
Money	0.005 (0.029)
Info	−0.118 (0.028)***
Percent Given	0.001 (0.009)
R^2	0.219
Adj. R^2	0.143
Num. obs.	114

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2: Ordinary least squares regression of donor designation rate at state-year level using objective form characteristics only. Dependent variable is designation rate as a fraction. Level of observation is the state-year. Year dummies included in regressions but excluded from table output. Standard errors clustered at the state level.

Incidentally, the model in Table 2 also suggests that the provision of information about the meaning of donor designation may actually *lower* donor designations rates. This result is also somewhat in contrast to Kessler and Roth (2014a), who include in their paper an experiment with real donation decisions where some subjects are exposed to a list of organs that may be used for donation. They find that this exposure increases the registration rate in their experiment. In contrast, my panel regression suggests that additional information can decrease the interest rate. However, the information provided on the DMV forms is quite different than the list of organs provided by Kessler and Roth, so again my results are not directly contradictory.

3.2 Non-choice Approach

To understand the motivation for this non-choice approach to this problem, consider what would be the optimal experimental design to determine the effect of choice framing. Ideally, each state would create two versions of their form: one opt-in and one active choice, but with all else held equal. These forms would be randomly assigned to DMV customers, and the designation rate could be compared for the two versions of the form. This approach would yield an estimate of the effect of choice framing for any state.

Using the non-choice approach, I simulate this ideal experiment design without having to implement a costly field experiment. To do this, I created a counter-factual version of each form in my data: I edited the form to change the apparent choice framing, but kept the form otherwise identical. For forms with an active choice framing, the negative option was removed. For forms

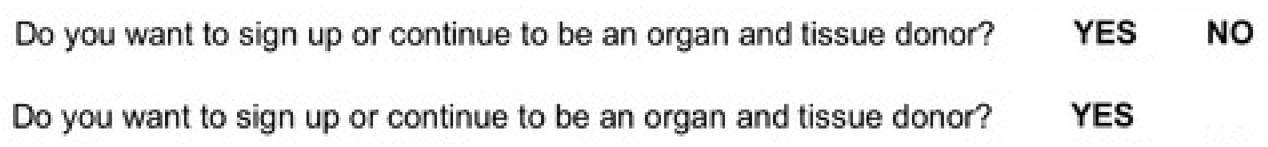


Figure 5: Example of a photomanipulation to switch the choice framing of a DMV form. The original wording of the quesiton used an actice choice framing (top). Simply removing the "No" option results in an opt-in choice framing (bottom).

with a passive or opt-in framing, a negative option was added. For an example of a factual and counter-factual pair of forms, see Figure 5. Importantly, the forms were also censored to make it difficult to tell which state each form was from; this reduced the chance that state-specific norms or beliefs could affect a subject's judgment of the forms.

3.2.1 Design

The non-choice responses were collected from participants on Amazon Mechanical Turk. Survey respondents were shown 10 randomly selected forms.[8] For each form, the respondent was asked to think about the typical person in their state applying for a driver's license with this form. They were then asked to give their responses to the questions laid out in Table 3. It was made clear to the respondents that they were not being judged or paid based on their accuracy. Participants were paid a fixed amount of $3.00 for completing the survey. A screenshot of a decision screen for this part of the study is shown in Figure B.1.

3.2.2 Data

Responses were collected from 572 Mechanical Turkers; 44 percent were female, and the majority had completed at least some college education. The median age was 30. Subjects took 25 minutes to complete the survey on average.

3.2.3 Model Selection and Prediction

I randomly divide the existing forms into 5 folds, and hold out one fold. The model selection algorithms (and the benchmark panel regression) are run on the remaining four folds, with the held out fold used to calculate prediction error. This process is repeated for each fold, so that a prediction is generated for each form-year. The prediction metrics are then calculated at the form-year level.

The results of this approach are in Table 4. I use all data from 2007 to 2012 where I have both the designation rate for that state-year and the application form for that state-year. This gives

[8] All forms are visible at `http://stanford.edu/~jnaecker/files/states`.

Questions	Responses
How likely do you think the typical person in the United States would be to become an organ donor when filling out this form?	• Very unlikely • Somewhat unlikely • Equally likely and unlikely • Somewhat likely • Very likely
How how hard was it for you to find the organ donation question?	• Very easy • Somewhat easy • Neutral • Somewhat hard • Very hard
How hard to read is the form?	• Very easy • Somewhat easy • Neutral • Somewhat hard • Very hard
To what extent do you think that the way this form asks about organ donation makes ... • an individual feel good about themselves by signing up? • an individual appear generous to others by signing up? • it easy for an individual to sign up? • an individual feel pressured to sign up?	• Not at all • Moderately • Somewhat • A fair amount • Very much

Table 3: Non-choice questions for the organ donation survey, along with allowed answers. All responses were on a five-point scale. All participants answered all questions in relation to 10 randomly selected forms.

Method	Mean Actual	Mean Pred.	Bias	MSPE	Calib.
Year Only	0.430	0.428	-0.002	0.023	-0.230
Applied	0.430	0.405	-0.025	0.044	-0.086
Likelihood ≥ 4	0.430	0.419	-0.011	0.023	-1.633
Likelihood ≥ 4 + Good ≥ 4	0.430	0.419	-0.011	0.021	0.499
Year + Likelihood ≥ 4	0.430	0.420	-0.010	0.023	-0.241
Year + Likelihood ≥ 4 + Good ≥ 4	0.430	0.418	-0.012	0.022	0.444
Lasso	0.430	0.410	-0.020	0.049	-0.387
Lasso (manual λ)	0.430	0.421	-0.009	0.025	-0.372
BIC Stepwise	0.430	0.457	0.028	0.223	-0.063
BIC Stepwise (Restricted)	0.430	0.395	-0.034	0.031	-0.024
CV-MSPE	0.430	0.400	-0.030	0.038	-0.461
CV-MSPE (Restricted)	0.430	0.413	-0.017	0.032	-0.627
Lasso (Year)	0.430	0.416	-0.014	0.030	-0.143
Lasso (Year, manual λ)	0.430	0.420	-0.010	0.026	-0.096
BIC Stepwise (Year)	0.430	0.316	-0.113	0.127	-0.108
BIC Stepwise (Restricted, Year)	0.430	0.398	-0.031	0.035	-0.119
CV-MSPE (Year)	0.430	0.390	-0.040	0.049	-0.182
CV-MSPE (Restricted, Year)	0.430	0.417	-0.013	0.035	-0.470

Table 4: Out-of-sample performance of non-choice responses compared to various objective data benchmarks. The first section of the table shows the performance of the benchmarks that use only objective data as predictors. The next sections shows the best-performing univartiate and bivariate non-choice data models. The third section shows the performance of these models with year fixed effects added. The fourth section shows the performance of the three model selection procedures, both with and without additional tightening of the over-fitting penalties. The final section shows the performance of these same model selection methods with year fixed effects added. Estimation is done using five randomly selected folds. Data are from 2007 to 2012.

114 observations over 6 years, with 29 different states appearing at some point in the unbalanced panel.

In the first section of Table 4, I include the prediction statistics for the two benchmarks models. The "Applied Model" benchmark runs the OLS regression from Table 2 on the training data, and uses the resulting model to predict the designation rates held-out forms. I also include benchmark model that uses only the year dummies. For our non-choice models to be useful, they should have better prediction statistics than these two benchmarks.

Before considering the model selection methods, I examine first some manually selected non-choice models that use a small number of non-choice variables. In the second section of the table, we see that predictions using only the hypothetical likelihood question have a low bias and MSPE, but extremely poor calibration. Adding a subjective dimension, however, can improve prediction dramatically. The best improvement from a single variable comes from adding the a subjective predictor derived from the question "To what extent do you think that the way this form asks about organ donation makes an individual feel good about themselves by signing up?"

In the third section of the table, I add a time trend to these two simple non-choice models. The predictive metrics are improved only slightly, indicating that the additional predictive power of the time trend is marginal when the key non-choice dimensions have been included. This is consistent with the fact that the average designation of the states increases only 2 percent per year, while the variance across states each year is much higher. Thus the simple non-choice models seem to be explaining a large portion of this cross-state variation.

Can we make better predictions by including more non-choice variables? Recall the intuition that we are attempting to capture the subjective representation of the forms, which possibly vary on many of these subjective dimensions. Intuitively, the hypothetical question may be a biased prediction of the true designation rate, and that this bias changes with an important subjective dimension of the forms. However, the additional non-choice questions may mediate this bias, allowing us to make more accurate predictions if we include these variables. To this end, the fourth section of Table 4 examines the predictive accuracy of the three main model selection procedures.

The Lasso algorithm, as discussed earlier in the paper, makes its selection of variables through the addition of an explicit penalty term proportional to the amount of coefficient in the model. As we can see from Table A.2, this algorithm fails to do much selection at all, despite the cross-validation procedure used to pick a penalty term λ. As a more conservative model selection procedure, the Lasso algorithm was run using a manually selected value for λ, chosen to result in models with approximately 4 degrees of freedom. This restriction leads to a full-sample model that is much sparser, and the MSPE is much improved. The BIC-selected and MSPE-selected models benefit similarly from a manual restriction. In these cases, the selection methods were forced to consider only models that included four degrees of freedom. All of these models select some function of the hypothetical likelihood response, and all but the un-restricted BIC-selected model are in fact identical.

In the final section of the table, I add a time trend to the set of variables that each model selection technique is allowed to consider, and force the techniques to include this variable.[9] The resulting predictive power of each selection technique is generally improved by the inclusion of the time trend. Because the reported prediction statistics are calculated from predictions made my models that had not seen that data point, the inclusion of additional variable does not necessarily improve the performance metrics. As above, it appears that the non-choice models can explain much of the inter-state variation in designation rates, which is much large in magnitude than the time trend.

[9] Note that time is included as a single trend variable and not as a set of year dummies as in the applied specification.

Method	Effect
Lasso	-0.019
Lasso (manual λ)	-0.025
BIC-selected	0.025
BIC-selected (restricted)	-0.021
RMSE-selected	-0.021
RMSE-selected (restricted)	-0.021

Table 5: Average predicted effect of changing from opt-in framing to active choice framing for each model selection procedure.

3.2.4 Effect of Choice Framing

We can now examine what our non-choice models have to say about the effect of choice framing on the designation rate. Recall that for each existing form, I created a counterfactual version with the opposite choice framing. I collected non-choice responses to both versions, so I am able to construct the the following object:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}(X^A - X^O),$$

where i indexes the nn forms, $\hat{\beta}$ is the non-choice model estimated using the existing forms, X^A contains the non-choice variables corresponding to the active choice forms (both factual and counterfactual), and X^O contains the non-choice data from the opt-in forms. Thus the object $\hat{a}$ is the average predicted effect of changing from opt-in to active choice framing. In Table 5, I give these counterfactual predictions for each of the non-choice model selection techniques. From this table, we can see that the predicted effect is negative in sign and around 2 percentage point in magnitude for nearly every non-choice selection procedure. The model that is selected by three of the methods gives a prediction of a negative 2.1 percentage point decrease from switching to the active choice framing. This point estimate is nearly identical to that of Kessler and Roth (2014a).

4 Social Preference Experiment

The non-choice response approach in Bernheim et al. (2013) has been validated so far only in relatively simple individual decision-making problems in the laboratory. However, the organ donation application (and other public good applications) contains at least two complicating features relative to these experiments. First, the application will use a new population of respondents (Mechanical Turk workers) to predict behavior. Second, the decision to donate contains many more potential subjective factors, such as social image and morality, that were not present in previous studies. To examine the potential effect of these issues, I will now apply the non-choice methodology to a set of social preference experiments on Mechanical Turk. The results of this experiment will help us determine whether the new methodology can work on a non-laboratory population and whether it can capture key determinants of more complicated social decisions. The experiment serves an additional purpose: I can estimate the non-choice model in one economic setting (a simple dictator game) and compare its predictions to actual behavior in another settings (a more complicate game). Furthermore, I can use a variety of more standard approaches (both structural and reduced form) as benchmarks to gauge the efficacy of the non-choice approach.

4.1 Design

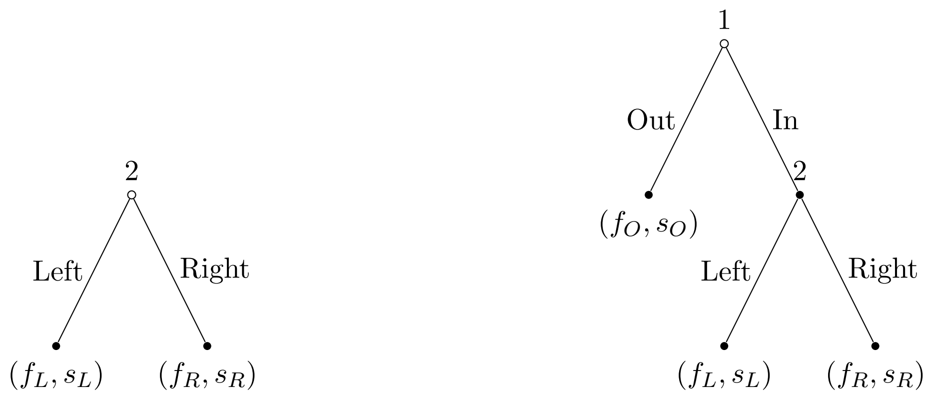
The experiment consists of two distinct sources of data collected from different groups of participants. First, I will discuss the choice data of interest, which come from to simple laboratory games. The participants in these games made binary choices that affected their own payoffs and others, much like the decision to donate one's organs. I will then go over how the non-choice responses about these two games were collected.

4.1.1 Choice Data

Subjects in the choice part of the experiment played two different binary-choice games: one in which both players make a move, and one in which only one player was able to act. I will first discuss the two-mover game, as the one-move game follows from this design.

Two-mover game. Player 1 moves first, and chooses either "Out" or "In". If "Out" is chosen, the game ends; Player 1 receives payoff of f_O in dollars and Player 2 receives payoff s_O in dollars. If "In" is chosen, Player 2 may choose either "Left" or "Right". If she chooses "Left", the game ends and payoffs are f_L and s_L respectively. If she chooses "Right", the game ends with payoffs f_R and s_R. An example of such a game is shown in Figure 6. The design of this game is based on that of Ert et al. (2011), which in turn is similar to the design used in Charness and Rabin (2002).

In Figure 7, I plot several example versions of the two-mover game, which differ by the payoff parameters $(f_O, f_L, f_R, s_O, s_L, s_R)$. The first mover's payoff is plotted on the horizontal axis. Similarly, the second mover's payoff is plotted on the vertical axis. Thus each point on the graph corresponds to a terminal node of the game; the points are labeled with the corresponding actions preceding them. We can see that Player 1 must decide between ending the game with "Out" or continuing the game with "In". If Player 2 gets to make a choice, she chooses between "Left" and "Right", the payoffs for which I have connected with a line. The slope of this line determines what kind of trade-off (if any) Player 2 must make in her decision. If the slope is negative, Player 2 must give up some of her payoffs to increase the payoffs of Player 1. If the slope is very steep, then giving is very expensive: she must give up more than one unit of her payoffs to increase the other's



(a) The one-mover game, which is a binary dictator game. Note this game is simply the proper subgame of the two-mover game.

(b) The two-mover game game. Player 1 moves first, choosing "Out" or "In". Player 2 then chooses "Left" or "Right".

Figure 6: Binary choice games used in binary choice experiment. The payoffs f are those for player 1, while the payoffs s are for player 2.

payoffs by one unit. If the slope is very shallow, giving is very cheap: she needs to give up less than one unit of her payoffs to increase the other's payoffs by one unit. In some cases, the slope may be positive or zero, in which case the players agree on which is the best outcome between "Left" and "Right" in terms of monetary payoffs. Nonetheless, it is possible that Player 2 will pick the dominated outcome between the two, perhaps to punish or spite Player 1.

Participants played 120 versions of the game with different payoff amounts $(f_O, f_L, f_R, s_O, s_L, s_R)$. These 120 sets were the same as those chosen by Ert et al. (2011) to cover a large space of possible social preference games often used in the experimental literature.[10] For example, if $f_R > f_O > f_L$ and $s_L > s_R > s_O$, this is the so-called "trust game": the first mover can make both players better off by playing "In", but the second mover can take these gains for herself by choosing the more selfish of her two options. (See version number 44 in Figure 7 for an example of such a game.) If $f_R > f_O > f_L$ and $s_O > s_R > s_L$, the situation is similar to that of the "ultimatum game": the first mover can play selfishly by choosing "In", but the second mover can then punish him at some cost to herself. (See version number 86 in Figure 7.) By including such a broad set of payoffs, we can be more confident that any results found are a feature of the entire class and not just a small corner of the payoff space.

One-mover game. In this game, Player 1 does not move. Player 2 chooses between "Left", with payoffs f_L and s_L, and "Right", for payoffs f_R and s_R. An example of this game is shown in Figure 6. This is a binary version of the standard dictator game used in numerous social preference experiments. Importantly, there are 120 versions of this game, each one corresponding to the subgame of one of the versions of the 2-mover game. That is, subjects playing the 1-mover game

[10]See Table A.1 for the complete list of games and their payoffs.

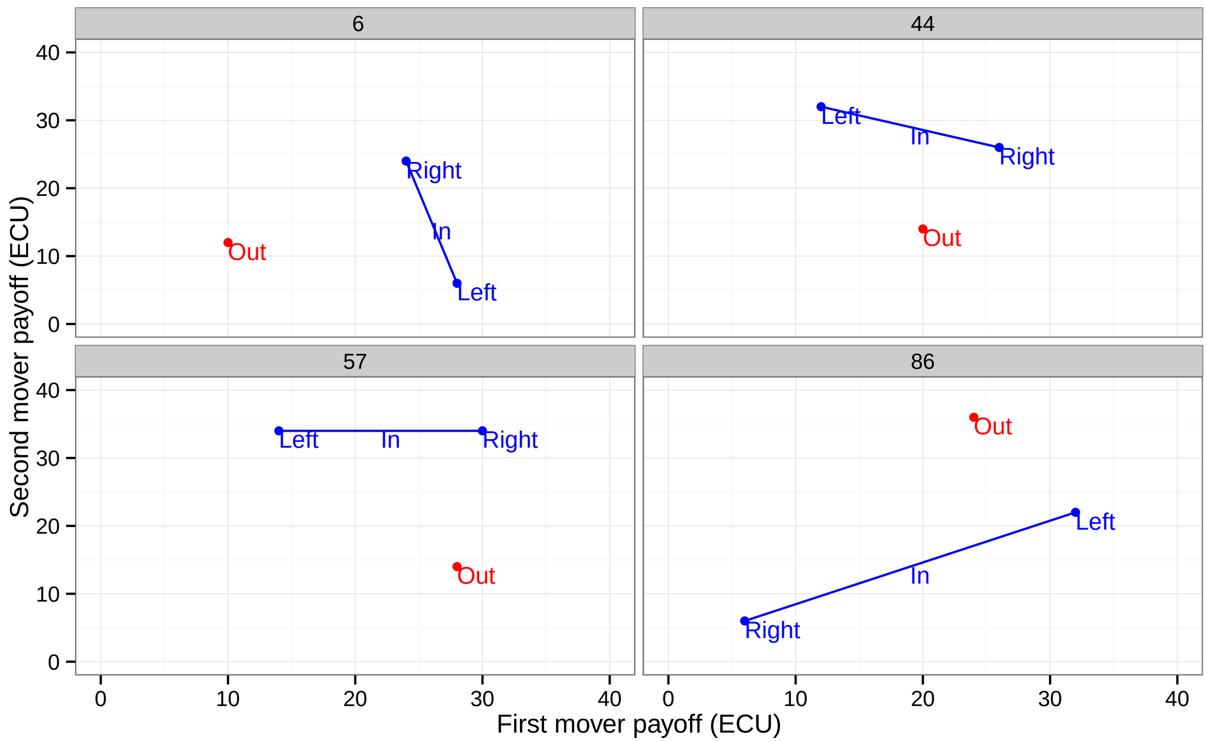


Figure 7: Payoffs from four selected versions of the two-mover game. Payoffs for first mover on horizonal axis, and payoffs for second mover on vertical axis. Version numbers 6 and 44 show a negative slope for Player 2's budget set, indicating that giving is costly. Version 57 shows a case where giving is costless, while number 86 shows a case where Player 2 has a an option that Pareto-dominates the other in terms of cash payoffs. Note that version 44 is an example of a simplified trust game: the first mover can move "In" in the hopes that the second mover will choose to make a small sacrifice to choose "Right." Version 86 shares some attributes of the classic ultimatum game: the second mover can "reject" the first mover's proposal of ending at the "Left" node by playing "Right."

face the same set of payoffs as in the 2-mover game, but without the context of their opponents choosing "In" or "Out" first. To emphasize this fact, I have labeled the dictator in the one-mover game as Player 2, and noted her payoffs with the same symbols following the "Left" and "Right" terminal nodes for the two games. This feature of the experiment was not made salient to participants, however.

Implementation details. The experiment was run on Amazon Mechanical Turk ("MTurk"). This is an online labor market where workers ("Turkers") perform short, computer-based tasks for modest wages (typically around $5 per hour). The platform was originally intended for businesses to be able to "automate" repetitive digital tasks, such as transcription of audio or tagging of image content. However, recently many social scientists have used the platform to perform large-scale laboratory-style experiments. The population of Turkers is not representative of the population in the United States, as Turkers skew younger and are more likely to be male.[11] However, Peysakhovich et al. (2014) show that the behavior of Turkers is generally similar to that of standard laboratory subjects in a wide variety of social preferences tasks, and that Turkers demonstrate individual consistency across these tasks and across time.

Subjects played all 120 versions of the two-mover game as both the first mover and the second mover, as well as the 120 version of the one-mover game, for total of 360 decisions. The payoffs f and s were displayed to subjects in experimental currency units (ECU), such that 10 ECU = \$1. The payoffs f and s were between 0 and 40 ECU (between \$0 and \$4) The experiment was implemented using the strategy method. That is, subjects made all 360 choices without being explicitly matched to a partner. When playing as the second mover in the trust game, the subjects made their actions conditional on their opponent choosing "In". After making all of their decisions, the players were matched in pairs, roles were randomly allocated, and one round of one game was implemented. All subjects received \$5 for participation automatically upon completion of the experiment, transmitted electronically through Mechanical Turk's payment system to the Turker's bank account. Any winnings from the implemented round (which averaged \$2) were calculated after all subjects had completed, and transmitted as a second payment several days later.

4.1.2 Non-choice Data

Four additional groups of subjects were tasked with providing non-choice responses to each decision problem. These subjects were shown the same instructions and games as the choice data subjects, but they were told that instead of making a choice, they would give unincentivized responses about the choice problems. Subjects were randomly assigned to one of four treatments, each of which asked a different non-choice question or set of questions. (See Figure B.2 for the exact wording of the instructions for these treatments.) The subjects in these four treatments answered the same non-choice question or questions for their entire participation in the experiment.

Subjects in the hypothetical question treatment answered the following question: "Hypothetically, which option would you choose?" Allowed responses were "Left" or "Right" when asked about their actions for the roles of Player 2, and "Out" or "In" when asked about the role of Player 1. Subjects in the vicarious hypothetical treatment answered the following question: "Which option do you think the typical Mechanical Turker would choose?" Allowed responses were "Left" or "Right" when asked about the roles of Player 2, and "Out" or "In" when asked about the role of Player 1. The vicarious likelihood question is motivated similarly, but worded so that participants can give a responses on a 5-point likelihood scale.

Finally, the subjective treatment asked participants about four possible dimensions of the subjective representation of these binary games. All the questions of this type were motivated by

[11] Turkers are required by Amazon to have American bank accounts. Some English-speaking foreigners are able to enter the labor force nonetheless, as determined by looking at their IP addresses. In my data, approximately 10% of subjects appear to be foreign.

leading theories of social preferences. The question about "feeling good" was intended to capture the effect of warm glow preferences, for example. These questions were worded so as to ask how these motivations addressed the concerns of a typical Mechanical Turker. As with the likelihood treatment, these questions allowed for response at one of five levels.

Implementation Details. Participants in the non-choice response treatments answered their question or questions for 360 unique settings: the 120 versions of the 1-mover game and 120 versions of both perspectives in the two-mover game game. Because these were non-choice responses, payments did not depend on the responses of the subjects. Instead, subjects in these treatments were paid a fixed amount of \$7. This payment arrived as a \$5 completion payment made immediately as in the choice treatment, followed by a \$2 payment several days later to match the timing of the choice treatments as much as possible.

4.2 Results

4.2.1 Overview of Data

The data for this experiment was collected from 149 subjects on Mechanical Turk in June of 2014. The median age of subjects was 29, and they were 45% female. Despite what one might expect given the low wages offered on the platform, 85% of subjects reported at least some college education. Subject were randomly assigned to treatments upon starting the experiment; the number of subjects in each treatment is given in Table 7. The median time to complete the experiment was 53 minutes across all treatments.

Before moving on, it is important to note that the following analysis will focus only on the behavior of subjects in the one-mover game and as the second mover in the two-mover game. By examining only the behavior of the final move in each game, we do not have to worry about beliefs of the subjects about what the second-movers will do. By design, each decision in the dictator game has a payoff-equivalent decision for the second mover of the two-mover game. This will allow us to directly examine whether final payoffs alone enter preferences, or whether subjects consider the path taken those payoffs as well. When predicting behavior in the two-mover game from observed behavior in the one-mover game, we now have a benchmark by which to judge our predictions: if only final payoffs matter, then the subjects should make the same decisions in matched versions of the two games.

4.2.2 Non-Parametric Results

We start by examining how behavior in the two games changes as the parameters of the choice problems change. The most salient feature of the decision problems that the final mover in each game faces is the relative payoffs of the two players. To summarize the relative payoffs for each

Treatment	Question(s)	Responses
1. Hypothetical	Hypothetically, which option would you choose?	• Left • Right
2. Vicarious Hypothetical	Which option do you think the typical Mechanical Turker would choose?	• Left • Right
3. Vicarious Likelihood	Which option do you think the typical Mechanical Turker would be more likely to choose?	• Very likely to choose Left • Somewhat more likely to choose Left • Equally likely to choose either option • Somewhat more likely to choose Right • Very likely to choose Right
4. Subjective Questions	Which option do you think more effectively addresses a typical Mechanical Turker's concern about . . . • feeling good about themselves? • appearing generous to others? • how their payoff will compare to others? • how much money the other player deserves?	• Left much more effective • Left somewhat more effective • Left and Right equally effective • Right somewhat more effective • Right much more effective

Table 6: Non-choice treatments and their corresponding questions and responses. Treatments were across subjects, meaning that a subject answered the same question or questions for the entire length of the experiment. For the first player perspective in the two-player game, "Left" and "Right" were replaced with "Out" and "In", respectively.

Treatment	Number of Subjects
Real Choices	36
Hypothetical	30
Vicarious Hypothetical	33
Vicarious Likelihood	33
Subjective Questions	17

Table 7: Number of subjects in each treatment. Mechanical Turkers were randomly assigned to treatment upon starting the experiment. The number of subjects in the subjective questions treatment is noticeably smaller due to attrition. This treatment took somewhat longer to complete, which may have caused more subjects to drop out or fail to complete in the time limit given. If such selection is present, it should not affect the performance of the non-choice methodology, as the subjects in the subjective questions treatment gave responses to both the one mover and two mover games.

version, I define the price of giving as

$$p = log\left(-\frac{f_L - s_L}{f_R - s_R}\right).$$

If the price is greater than zero, then Player 2 must sacrifice more than one dollar to give a dollar to their partner. Conversely, if the price is less than zero, the decision maker needs to give up less than one dollar to give a dollar to their partner. Of course, since the decision in this experiment is binary, subjects can make only extensive-margin decisions about whether or not to choose the selfish option rather than an intensive-margin decision about how much to give. So, we should expect that the proportion of individuals choosing the selfish option should increase with the price of giving. Note that this price is only well-defined for cases where the selfish option for the decision maker is not also the preferred option of their partner. These other cases, where incentives of the two players are aligned or where one or both player's payoffs do not depend on the choices of Player 2, are dropped from this part of the analysis.

Figure 8 plots the fraction of subjects choosing the selfish option against the price of giving. As expected, subjects choose the selfish option more often on average as the price of giving increases. The figure plots the average demand separately for the two games. If final payoffs were the only variables than entered the subjects' utility functions, we should see no difference between the two games. However, we see than the demand for the selfish option depends noticeably on which game is being played. In particulars, subjects make consistently less selfish choices at every price level in the two-mover game. These results are confirmed by a regression analysis: in Table 8, I report the results of probit regression of individual binary choices on price of giving and an indicator for game type. Model 1 confirms that in aggregate, demand of the selfish option increases with price. Model 2 shows that the two mover game has a lower level of giving, though not a significantly different slope. This gives us our first hint that preferences do not depend solely on monetary outcomes:

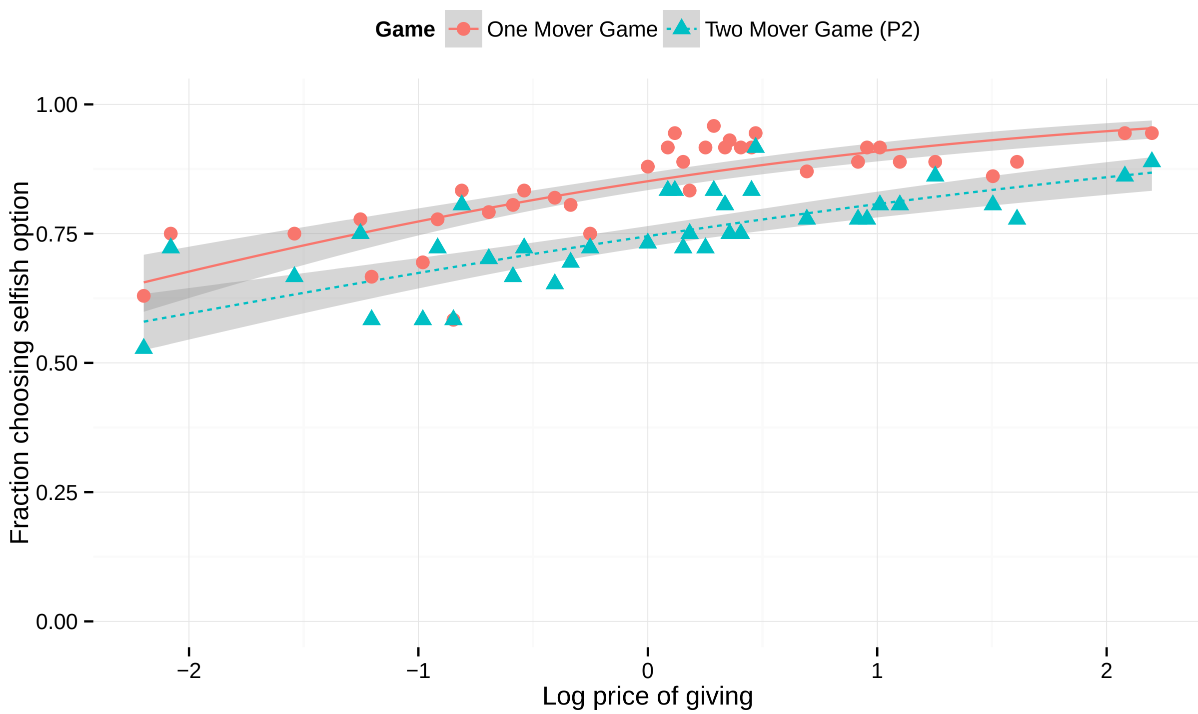


Figure 8: Fraction of participants choosing selfish option for each game, as a function of the log price of giving. The points correspond to one version of each game. Also plotted are probit response curves for individual demand by Player 2 for the selfish option as a function of the price of giving, plotted separately for the one-mover and two-mover games. Data is only from games where slope of Player 2's budget set is negative. Individual data points are not shown because all points lie at level either 0 or 1 on the vertical axis. Only games with a well-defined log price are shown.

clearly, whether those outcomes are part of the two-player or one-mover game have a significant effect on choices. To understand the difference between the two games, we turn to a structural model of social preferences.

4.2.3 Benchmark Models of Social Preferences

In order to judge how useful the non-choice responses are in prediction choice behavior, I discuss in this section several benchmark approaches that use choice data to predict behavior in a social preference setting. The results below show that the subjects in this study have preferences that are very similar to those of subjects in previous papers in the literature. However, I will also demonstrate how their behavior shows a dependence on context that is not captured by the benchmark approaches.

	Model 1	Model 2
Constant	0.833***	1.043***
	(0.136)	(0.169)
Price	0.240***	0.292***
	(0.052)	(0.073)
Two-mover		−0.383*
		(0.197)
Price X two-mover		−0.084
		(0.063)
AIC	3736.461	3676.477
BIC	3748.916	3701.389
Log Likelihood	-1866.230	-1834.238
Deviance	3732.461	3668.477
Num. obs.	3744	3744

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 8: Probit regressions of individual choices on price of giving and game type. Dependent variable is indicator for whether subject choose selfish option. Standard errrors clustered at the subject level.

Structural Model. The first approach with which I will benchmark the non-choice prediction is a leading structural model of social preferences. I will use the model proposed by Charness and Rabin (2002). In this model, we imagine a setting where money is being divided between two agents, 1 and 2. Agent 1's utility depends on x_1, the dollar amount that she receives, and x_2, the dollar amount her partner receives. Utility for agent 1 takes the following form:

$$u(x_1, x_2) = (1 - \rho r - \sigma s - \theta q)x_1 + (\rho r + \sigma s + \theta q)x_2,$$

The terms ρ, σ, and θ are parameters of the utility function, while the terms r, s, and q are determined by the situation. We can see that the utility function puts weights on each of the dollar payoffs, x_1 and x_2. These weights always sum to one, but their relative size varies depending on the situation. First, we define the distributional terms r and s. If $x_1 > x_2$, then we set $r = 1$ and $s = 0$. Thus ρ is the weight that the decision-maker puts on their partner's payoff when the decision-maker is ahead in terms of payoff. If $x_1 < x_2$, then set $s = 1$ and $r = 0$. Thus σ is the weight that the decision-maker puts on their partner's payoff when the partner is ahead in terms of payoff. Finally, we define $q = -1$ if the partner has "misbehaved", and is 0 otherwise. "Misbehaved" in the game analyzed in this section is defined as first mover choosing "In" when "Out" would have been best joint payoff *and* best payoff for second mover.[12] Thus θ measures the

[12]This is the definition used by Charness and Rabin (2002) as well as Ert et al. (2011), from whom the different

reciprocal nature of the decision-maker's preferences.

This model provides a useful benchmark because it incorporates a wide variety of commonly-used social preference models. For example, if a subject cares only about her own payoff, then we should find $\rho = \sigma = \theta = 0$. If instead she cares only about maximizing total payoffs (ie utilitarian preferences), then we should find $\rho = \sigma = \frac{1}{2}$ and $\theta = 0$. And if her utility is equal to the minimum payoff (Rawlsian preferences), we should see $\rho, \sigma \to \infty$ and $\theta = 0$. This formulation of the utility function also includes the model of Fehr and Schmidt (1999) as a special case where $\theta = 0$.

Note that if $\theta = 0$, the preferences represented are purely distributional: the final payoffs of each player enter in the utility function, but how those allocations were arrived at does not. It is possible, however, that preferences are affected by the payoffs of unrealized outcomes; that is, the *path* matters, not just the outcome. The Charness-Rabin model allows for this possibility in a fairly flexible way: the parameter θ indicates how much penalty the decision-maker applies to their opponent when the opponent wrongs her.[13] The importance of this reciprocity component can be seen in comparing the one-player and two-mover games in this study.

I estimate the Charness-Rabin model using a maximum likelihood procedure, with an extreme value error distribution. In this case, the probability of choosing option Right over option Left is given by

$$P(R) = \frac{e^{\mu U_R}}{e^{\mu U_L} + e^{\mu U_R}},$$

where U_R and U_L are the utility values of choosing Right and Left, respectively. The parameter μ governs the error rate: a higher value indicates a smaller chance of choosing the lower-utility option over the higher-utility one. The estimates I report assume a representative agent, ignoring the possibility of heterogeneous preferences across agents. While heterogeneity is a potentially interesting direction of study, I make this choice for two reasons. First, it allows me to judge the non-choice models against the structural models in the case where both have the same degrees of freedom. Second, the non-choice methodology I will employ will predict only population-level choice frequencies, so again to level the playing field I choose to limit the structural methods to the population level as well.

The results of this estimation are shown in Table 9. In the first two columns, I examine the results when aggregating data from the two games. In the first column, I restrict $\theta = 0$. We see that since $\rho > 0$ and $\sigma > 0$, subjects put significant weight on other's payoffs. However, the fact that $\rho > \sigma$ means that subjects on average put less weight on others' payoffs when the other is ahead in terms of dollar payoffs. In the second column, I allow θ to be determined by maximum likelihood. The negative estimate of θ indicates that subjects do indeed have a reciprocal components to their preferences. These point estimates are quite similar to those from Charness and Rabin (2002).

In the remaining columns, I consider the estimates we obtain when restricting the data to one

versions of each game are borrowed.

[13]One could also imaging more detailed models of reciprocity that explicitly model the expectations of the players.

	All data	All data	1 mover game	2 mover game	2 mover game
μ	0.2106 (0.0048)	0.211 (0.0048)	0.2379 (0.0077)	0.1924 (0.0063)	0.1928 (0.0063)
ρ	0.3397 (0.0145)	0.3391 (0.0145)	0.2809 (0.0193)	0.3977 (0.0215)	0.3966 (0.0215)
σ	0.1521 (0.015)	0.1396 (0.0159)	0.1171 (0.0202)	0.1871 (0.022)	0.1739 (0.0232)
θ	.	-0.0721 (0.0263)	.	.	-0.0744 (0.0369)

Table 9: Maximum likelihood estimates for the Charness-Rabin model under various restrictions. Errors are from extreme-value distribution, governed by parameter μ. Parameters ρ and θ measure distributional preferences, while θ measures reciprocal preferences. Model 1 estimates using second-mover choices from both games, but with the restriction that $heta = 0$. Model 2 estimates using the full data set as well, but with no restriction. Model 3 estimates CR using one-mover game data; in this case it is not possible to estimate $heta$, which is set to 0. Model 4 estimates CR using only data from the second mover of the two-mover game, with the restriction that $heta = 0$. Model 5 relaxes this restiction.

of the two games. If we restrict the reciprocity component $\theta = 0$, we see that we get very different estimates for ρ and σ across the two games (columns 3 and 4). Even if we drop the restriction $\theta = 0$ for the two-mover game, the two distributional parameters are different from those resulting from the one-mover game (column 5). Taken as a whole, these results shows us that while the Charness-Rabin model allows for a reciprocal term in the utility function, the model has failed to fully capture the differences in behavior between the one-mover and two-mover games. This opens the possibility that non-choice responses may be more useful as predictors than the Charness-Rabin estimates.

Heuristic Model. As another benchmark, we can instead model subjects arriving at decisions through simple heuristics, or rules of thumb, rather than utility maximization. There are a number of possible such rules, many of which correspond to the special cases of the Charness-Rabin model discussed above. However, these heuristics do not assume an underlying utility model, but simply specify a probability of choosing an option as a function of its characteristics. The heuristics rules that I will use are as follows:

- *Selfish.* The subject chooses the option with higher payoff for themselves.
- *Nice selfish.* The subject chooses the option with the higher payoff for themselves. If their own payoff is the same for the two options, they choose the option with higher payoff for their partner.
- *Utilitarian.* The subject chooses the option with the higher aggregate payoff for the two players.
- *Rawlsian.* The subject chooses the option with the higher minimum payoff.
- *Minimizing differences.* The subject chooses the option which has the smaller difference between the two player's payoffs.

To see which of these heuristics are being used by subjects, I regress subject decisions on five indicators. The dependent variable is an equal to 1 if the subject chose the option "Right", while the independent variables are equal to 1 if the option "Right" is the option that corresponds with the named heuristic, equal to 0 if the option "Left" meets the heuristic, and equal to 0.5 if both options meet the decision rule's criteria. The univariate models show that the selfish and nice-selfish heuristics give the most explanatory power, but that all rules are correlated with choices to some degree. The negative coefficient on the minimizing differences rule indicates that on average subjects sought to maximize differences between payoffs, but this effect is much smaller in magnitude than the other rules. Combining all heuristics into one model, we find that the nice-selfish rule is the dominant decision factor, which is qualitatively similar to the results in Erev and Roth (1998), who use a this same heuristic regression on their data. However, if we restrict the regressions by game, we see that in the one-mover game, the selfish decision rule is much stronger than in the two-mover game. This shows that like the Charness-Rabin model, the heuristic model also does not fully capture differences in behavior in the two settings.

4.2.4 Prediction

In the preceding analysis we have looked only at in-sample fit of the benchmark models, but the non-choice methodology is most likely to be useful when predicting out of sample. For what follows, I will perform the following exercise: Using choice data only from the one mover game, I will predict aggregate behavior in the two-mover game. As a benchmark, I will estimate the Charness-Rabin model with the one mover game data, but I will not be able to identify the reciprocity component θ, since there is no sense of one's partner "misbehaving" when they cannot make a move in the game. As another benchmark, I will estimate the reduced-form model from Table 10 using just the one mover game data. We have already seen that this model demonstrates some flavor of stability: the relative strength of the decision rules seems highly game-dependent.

It is possible that non-choice responses can overcome the instability prevalent in these other approaches. The procedure for making predictions with non-choice responses starts with creating the non-choice variables. The non-choice responses – which are collected at the individual level – are first aggregated to the level of the choice problem. For the binary responses, the most obvious aggregation is the mean, i.e. the fraction of non-choice subjects who respond with the selfish option to the hypothetical and vicarious hypothetical questions. The responses on the 5-point scale can be aggregated by taking averages as well, though this destroys much distributional information. Additionally, one can create variables corresponding to the percent of non-choice subjects who respond with level 2 or greater, 3 or greater, and so on.

Table 11 reports the predictive performance of the non-choice models and several benchmarks. The first second contains the benchmarks. The first benchmark makes predictions from the Charness-Rabin model estimated on the one mover game data. Recall that this approach does

	All	All	All	All	All	All	1M game	2M game
Intercept	0.162***	0.145***	0.212***	0.213***	0.566***	0.089***	0.067***	0.112***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
Selfish	0.719***					0.128***	0.174***	0.082***
	(0.007)					(0.010)	(0.010)	(0.012)
Nice Selfish		0.711***				0.448***	0.477***	0.419***
		(0.006)				(0.010)	(0.010)	(0.011)
Utilitarian			0.561***			0.131***	0.125***	0.138***
			(0.006)			(0.004)	(0.004)	(0.004)
Rawlsian				0.569***		0.118***	0.086***	0.150***
				(0.005)		(0.004)	(0.003)	(0.006)
Minimize Difference					−0.095***	−0.011***	−0.015***	−0.006**
					(0.003)	(0.002)	(0.001)	(0.003)
R^2	0.449	0.484	0.291	0.284	0.008	0.513	0.575	0.457
Adj. R^2	0.449	0.484	0.291	0.284	0.008	0.512	0.574	0.456
Num. obs.	8640	8640	8640	8640	8640	8640	4320	4320

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 10: Ordinary least squares regressions of individual choices on various social preference heuristic rules, using second-mover data from both games. Standard errors in parentheses, clustered at subject level. Dependent variable=1 if chose option Right, =0 otherwise. Selfish=1 if option Right with higher payoff for self. Nice Selfish=1 if option Right has higher payoff for self or payoffs for self are equal and payoff for other player is higher for Right. Utilitarian=1 if sum of payoffs higher for option Right. Rawlsian=1 if minimum payoff (between) two players is higher for option Right. Minimize Differences = 1 if option Right has smaller difference between payoffs for two players. All variables take value 0 if instead option Left meets criteria, and 0.5 if both options meet criteria.

not allow for estimation of θ, so I set $\theta = 0$ to make predictions. We see that this approach yields a very small bias (less than one percentage point in average error) but a relatively poor MSPE and calibration score. For a much more challenging benchmark, I can also estimate the Charness-Rabin model on the data from *both* games.[14] This additional data does not improve the prediction metrics; this is likely because, as we saw earlier, the Charness-Rabin does not seem to full capture the differences in behavior between the two games.

Alternatively, we can predict that subjects will always demand the selfish option. This approach does particularly poorly because it gives a myopic prediction that does not depend on the aspects of the game at all. The heuristics method fares much better, with the lowest MSPE among the benchmark models. The final benchmark, called version match, predicts that the aggregate choice frequency in any version of the two mover game is the same as the aggregate choice frequency in the payoff-matching version of the one-mover-game.

The rest of the table shows the performance of various non-choice models. First, I examine the predictive power of the first-person hypothetical question. The label "raw" indicates that the variable has been included directly as a prediction of the real demand, without the regression step explained above. The raw hypothetical response turns out to be a relatively good predictor of choices; in this setting, there appears to be only moderate hypothetical bias. The "trained" model is only a marginally better predictor; this is due to the fact that the hypothetical bias is not perfectly stable across the two games. The question is whether the addition of subjective dimensions can further improve the predictive power of this hypothetical question.

In the next section of Table 11, I perform the full model selection procedures on all non-choice variables. For the models marked as "restricted", the search procedure did not consider models that had more than 4 variables (including the constant); this was done to limit these approaches to as many degrees of freedom as the Charness-Rabin benchmark has. From the table, it is clear that this restriction does not hurt predictive performance, and can actually improve it noticeably. The final section of the table given the non-choice models access to objective variables as well, in particular the payoffs of the "Left" and "Right" terminal nodes. Again, this has little effect on predictive performance of the various model selection techniques.

All of the resulting non-choice models have lower MSPEs than any of the benchmarks approaches, as well as reasonable biases and calibration scores. In fact, we can see in Table A.3 that the resulting models are very similar, and often identical. The hypothetical first-person response is included by all selection methods. The subjective asking about "appearing generous" and "how much money the other deserves" also consistently appear in the non-choice models, indicating that these variables play a key part in the subjective representation. In contrast, the variables derived from the questions about "how their payoff will compare to others" and "feeling good about themselves" appear rarely in the non-choice models, if at all. This result suggests that warm glow and

[14] Again, the non-choice approach will not use data from the two mover game for estimation.

Method	Mean Actual	Mean Pred.	Bias	MSPE	Calib.
Charness-Rabin	0.8328	0.8254	-0.0074	0.0178	0.4440
Charness-Rabin (all data)	0.8328	0.7963	-0.0365	0.0171	0.4828
Selfish	0.8328	1.0000	0.1672	0.0427	
Heuristics	0.8328	0.8862	0.0534	0.0081	1.3038
Version match	0.8328	0.8862	0.0534	0.0081	1.1094
Hypothetical Response (raw)	0.8328	0.8654	0.0326	0.0047	0.8087
Hypothetical Response (trained)	0.8328	0.8675	0.0347	0.0044	1.1637
Lasso	0.8328	0.8701	0.0373	0.0052	1.4148
BIC	0.8328	0.8675	0.0347	0.0044	1.1637
BIC (restricted)	0.8328	0.8675	0.0347	0.0044	1.1637
CV-MSPE	0.8328	0.8130	-0.0198	0.0079	0.6739
CV-MSPE (restricted)	0.8328	0.8689	0.0361	0.0041	1.1977
Lasso with objective	0.8328	0.8694	0.0366	0.0050	1.3923
BIC with objective	0.8328	0.8675	0.0347	0.0044	1.1637
BIC with objective (restricted)	0.8328	0.8675	0.0347	0.0044	1.1637
CV-MSPE with objective	0.8328	0.8216	-0.0112	0.0054	0.7772
CV-MSPE with objective (restricted)	0.8328	0.8689	0.0361	0.0041	1.1977

Table 11: Statistics for all methods predicting average behavior in the two-mover game. The first segment includes the benchmarks against which the non-choice models can be judged. The Charness-Rabin predictions are estimated using maximum likelihood. The (all) version generates its estimates from both the one-mover and two-mover game. The selfish benchmark assumes all choices made with pure self-interest. The heuristics benchmark uses the full regression model from table 10 run on the one-mover game only. The version match benchmark assumes that the choice frequency in each version of the two-mover game is the same as the choice frequency in the corresponding version of the one-mover game. The remaining segments document the predictive performance of models that use the non-choice responses. The calibration score is slope coefficient from regression of actual values on predicted.

status matter less for the subjective representation.

4.2.5 Counterfactuals

Ultimately, we are not interested just in being able to make accurate predictions, but in using these predictions to learn about behavior. As an example, suppose we only had data on behavior in the one-mover game, as well as non-choice responses for both games. What could we learn about behavior in the two-mover game? One possibility is that we could test whether the reciprocity component in the full Charness-Rabin model is non-zero. Doing so with only choice data is impossible, as discussed above. It is possible using non-choice responses, however.

The first column of Table 12 gives the Charness-Rabin estimates from the two mover game choice

	Actual data	Hypotheticals	Vicarious	Lasso	BIC	CV-MSPE
μ	0.1928 (0.0063)	0.241 (0.0084)	0.1534 (0.0056)	0.1855 (0.0065)	0.2045 (0.0069)	0.2111 (0.0071)
ρ	0.3966 (0.0215)	0.4039 (0.0206)	0.3377 (0.0255)	0.2377 (0.0249)	0.3433 (0.0226)	0.3026 (0.022)
σ	0.1739 (0.0232)	0.1199 (0.0228)	0.6869 (0.0258)	0.0521 (0.0284)	0.0415 (0.0267)	0.1094 (0.0248)
θ	-0.0744 (0.0369)	0.0076 (0.0414)	0.3101 (0.0496)	-0.0437 (0.0466)	-0.0115 (0.0441)	-0.0598 (0.0401)

Table 12: Estimates for the Charness-Rabin model with actual and synthetic data from the 2-mover game. The first column reports the true estimates using the incentivized responses of the choice data subjects. The next two columns report estimates derived from the responses of the hypothetical and vicarious hypothetical groups. The last three columns report results from the fitted values of three selected non-choice models. For these models, synthetic data was generated by assuming that the predicted frequency of choosing the selfish option was the same as the actual choice frequency.

data; we are interested in whether we can recover these estimates using the non-choice approach. The second column contains the estimates using raw hypothetical responses as if they were real choices. The hypothetical responses generate estimates for σ and ρ, the distributional components, that are fairly close to the real data estimates. However, the hypothetical choices suggest that the reciprocity component θ is not statistically different from zero. Put another way, the hypothetical responses themselves give no indication that reciprocity matters for decision-making. The third column uses the third-party hypothetical question instead, and the resulting estimates of Charness-Rabin are even farther from the real data.

The remaining columns are estimated from the predicted choice frequencies of the non-choice models. Specifically, I generated synthetic data where the number of subjects was the same as the choice data, and the aggregate choice frequency for each version was the same as the predicted value from the non-choice model. The estimates from the non-choice models all have the correct sign on the reciprocity component, as well as reasonably close estimates for the distributional components and the error term. The MSPE-selected model, which has the lowest bias and MSPE of all the non-choice models, also gets the magnitude nearly exactly correct.

5 Conclusion

I have shown how non-choice responses can be used to predict behavior in two donation settings: sharing money in the lab and signing up to be an organ donor outside the laboratory. In both cases, responses from a relatively parsimonious set of questions can provide more accurate predictions of choice behavior than benchmark methods that are commonly used in those settings. Furthermore, I have shown through a simple "pseudo-experiment" that an opt-in choice framing leads to higher donation rates then an active choice framing. Predicting the magnitude of this effect is difficult with observational data and beyond the scope of our current behavioral theories.

Of course, there are limitations to the non-choice approach. One must have a sufficiently rich set of non-choice questions, so as to capture the subjective representation of the choice problem. From these questions, one must be able to construct a relationship that provides stable prediction across policy changes. The survey population must be able to approximate the subjective impressions of a potentially very dissimilar population in a potentially very unfamiliar setting. Empirical verification and refinement of non-choice questions, through repeated application of this methodology, can help us understand how successful it can be in new settings.

The strength of the non-choice data comes from the fact that it does not attempt to model the relationship from objective data to choices, but rather from subjective responses to choices. This approach has the advantage of being able to make predictions about changes in policies even when no variation in the dimension of interest has been observed before. Because of this, the non-choice approach may be useful for performing policy pseudo-experiments that are too difficult to run in reality, perhaps because their are prohibitively expensive or repugnant. Non-choice data can help us learn much more about behavior in these settings that we would not otherwise find out through standard methods.

References

Alpízar, Francisco, Fredrik Carlsson, and Peter Martinsson, "Using Choice Experiments for Non-Market Valuation," *Economic Issues*, 2003, *8*, 83–110.

Ashlagi, Itai, David Gamarnik, Michael A. Rees, and Alvin E. Roth, "The Need for (long) Chains in Kidney Exchange," May 2012. working paper.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones, "Can Marginal Rates of Substitution Be Inferred From Happiness Data? Evidence from Residency Choices," *American Economic Review*, October 2014, *104* (11), 3498–3528.

Bernheim, B. Douglas, Daniel Bjorkegren, Jeffrey K Naecker, and Antonio Rangel, "Do Hypothetical Choices and Non-Choice Ratings Reveal Preferences?," July 2013. working paper.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz, "Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research," *Journal of Marketing*, 2005, *69* (2), 1–14.

Charness, Gary and Matthew Rabin, "Understanding Social Preferences With Simple Tests," *The Quarterly Journal of Economics*, August 2002, *117* (3), 817–869.

Coffman, Lucas C., Clayton R. Featherstone, and Judd B. Kessler, "Can a Social Information Nudge Affect Job Choice? Experimental Evidence from Teach for America," 2013. working paper.

Erev, Ido and Alvin E Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, September 1998, *88* (4), 848–81.

Ert, Eyal, Ido Erev, and Alvin E. Roth, "A Choice Prediction Competition for Social Preferences in Simple Extensive Form Games: An Introduction," *Games*, July 2011, *2* (3), 257–276.

Fehr, Ernst and Klaus M. Schmidt, "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, August 1999, *114* (3), 817–868.

Jackman, Simon, "Correcting surveys for non-response and measurement error using auxiliary information," *Electoral Studies*, 1999, *18* (1), 7 – 27.

Kang, Min Jeong, Antonio Rangel, Mickael Camus, and Colin F. Camerer, "Hypothetical and Real Choice Differentially Activate Common Valuation Areas," *The Journal of Neuroscience*, 2011, *31* (2), 461–468.

Katz, Jonathan N. and Gabriel Katz, "Correcting for Survey Misreports Using Auxiliary Information with an Application to Estimating Turnout," *American Journal of Political Science*, 2010, *54* (3), 815–835.

Kessler, Judd B. and Alvin E. Roth, "Organ Allocation Policy and the Decision to Donate," *American Economic Review*, 2012, *102* (5), 2018–47.

_ **and** _ , "Don't Take 'No' For An Answer: An experiment with actual organ donor registrations," Working Paper 20378, National Bureau of Economic Research August 2014. working paper.

_ **and** _ , "Getting More Organs for Transplantation," *American Economic Review*, 2014, *104* (5), 425–30.

Kurz, Mordecai, "Experimental approach to the determination of the demand for public goods," *Journal of Public Economics*, 1974, *3* (4), 329 – 348.

Lebreton, Maël, Soledad Jorge, Vincent Michel, Bertrand Thirion, and Mathias Pessiglione, "An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging," *Neuron*, 2009, *64* (3), 431 – 439.

List, John A. and Craig Gallet, "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?," *Environmental & Resource Economics*, November 2001, *20* (3), 241–254.

Network, Website Organ Procurement Transplantation, "Organ Procurement and Transplantation Network," August 2014. website.

Peysakhovich, Alexander, Martin A. Nowak, and David G. Rand, "The Cooperative Phenotype," 2014. working paper - EMBARGO.

Rothschild, David and Justin Wolfers, "Forecasting Elections: Voter Intentions versus Expectations," March 2011. working paper.

Shogren, Jason, "Valuation in the Lab," *Environmental & Resource Economics*, 05 2006, *34* (1), 163–172.

Smith, Alec, B. Douglas Bernheim, Colin Camerer, and Antonio Rangel, "Neural Activity Reveals Preferences Without Choices," *American Economic Journal: Microeconomics*, April 2014, *6* (2), 1–36.

Thaler, Richard H, "Opting in vs. Opting Out," 2009.

Zou, Hui and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (2), 301–320.

Appendices

A Additional Tables and Figures

Version	f1	s1	f2	s2	f3	s3
1	14	6	6	30	14	32
2	34	8	30	26	12	20
3	6	12	10	36	10	36
4	14	12	18	28	26	34
5	16	6	32	30	16	36
6	18	28	6	26	34	20
7	14	18	22	6	12	32
8	20	24	10	22	24	22
9	22	6	12	28	26	28
10	16	30	12	34	16	16
11	16	22	30	6	30	6
12	24	6	26	26	28	8
13	30	8	20	22	22	6

14	8	24	16	12	32	10
15	22	12	22	24	32	30
16	20	36	22	8	8	26
17	20	20	30	12	10	10
18	26	12	16	36	26	10
19	22	8	28	30	18	8
20	20	14	12	32	26	26
21	8	24	20	6	36	6
22	10	34	8	16	36	8
23	14	10	26	14	32	18
24	24	12	14	30	26	16
25	6	22	8	22	20	26
26	18	28	22	20	30	16
27	28	22	8	22	34	30
28	30	6	10	6	36	20
29	20	22	16	30	30	10
30	24	4	34	18	32	6
31	8	8	24	14	22	24
32	14	32	24	12	30	14
33	28	20	4	4	30	30
34	24	24	26	6	26	6
35	20	22	32	22	36	8
36	10	26	34	6	12	30
37	8	22	16	8	22	14
38	24	36	32	22	6	6
39	30	34	34	6	20	28
40	28	28	34	24	14	22
41	22	34	26	24	12	8
42	28	14	14	34	30	34
43	6	22	32	28	20	8
44	6	32	10	24	12	34

45	18	34	28	32	16	30
46	12	16	34	32	12	26
47	34	6	8	8	12	8
48	34	6	22	10	30	24
49	30	16	28	10	14	30
50	18	24	20	22	10	14
51	24	6	18	26	34	8
52	36	22	20	26	24	24
53	8	24	6	16	20	24
54	18	26	34	22	16	36
55	32	20	10	28	20	16
56	20	16	36	26	34	6
57	26	20	12	26	16	16
58	8	20	18	10	8	6
59	8	30	24	20	34	4
60	4	26	32	28	14	16
61	20	24	18	32	24	32
62	24	8	26	16	12	30
63	30	26	14	32	20	36
64	32	16	22	22	28	14
65	6	12	12	10	8	10
66	8	16	22	12	12	12
67	14	18	10	14	32	20
68	18	28	24	12	14	16
69	12	36	14	28	22	12
70	8	8	20	36	18	8
71	34	18	30	28	10	34
72	20	20	34	20	10	34
73	30	14	24	20	28	30
74	14	34	20	10	14	8
75	10	26	8	30	20	22

76	34	6	30	32	6	10
77	14	24	20	10	8	34
78	32	18	10	32	36	36
79	20	20	10	30	18	8
80	18	16	32	20	14	22
81	8	10	20	18	12	8
82	20	36	32	30	36	26
83	8	8	18	18	30	10
84	12	22	22	4	36	20
85	34	34	18	6	36	18
86	10	22	20	8	12	12
87	26	16	24	20	32	26
88	10	32	12	20	12	20
89	32	6	14	14	32	6
90	12	12	20	6	34	12
91	24	8	6	18	20	8
92	20	26	30	16	20	30
93	24	28	6	34	26	10
94	26	8	26	8	32	10
95	22	26	30	24	12	26
96	24	32	14	32	26	14
97	20	32	26	6	32	16
98	28	36	32	26	22	24
99	10	34	16	18	28	12
100	32	18	6	10	30	28
101	24	8	34	4	32	8
102	18	8	12	16	12	14
103	6	10	10	30	8	14
104	10	12	28	6	24	24
105	36	10	8	10	14	32
106	28	18	32	26	20	16

107	28	14	30	6	24	14
108	34	4	28	20	24	24
109	22	16	32	32	14	34
110	22	34	20	24	26	24
111	10	32	28	12	30	28
112	30	8	22	18	32	18
113	28	6	24	12	34	12
114	12	22	14	32	32	22
115	26	24	6	8	30	14
116	22	22	24	18	24	18
117	14	8	36	20	34	26
118	10	8	34	34	28	8
119	24	12	6	8	24	14
120	16	6	14	8	20	26

Table A.1: List of all games used in the binary game section of the paper.

B Instructions and Decision Screens for Laboratory Experiment

Variable	Best Bivariate	Lasso	Lasso (manual λ)	BIC	BIC (Restricted)	MSPE	MSPE (Restricted)
(Intercept)	0.499	-0.936	0.506	0.409	0.499	-0.140	-0.024
Likelihood				0.175			
Likelihood ≥ 2							
Likelihood ≥ 3		-0.462					
Likelihood ≥ 4	0.481	0.029			0.481		
Likelihood ≥ 5							
Good							
Good ≥ 2							
Good ≥ 3		1.008					
Good ≥ 4	-0.871	-1.170	-0.029		-0.871		
Good ≥ 5			-0.260				
Generous							
Generous ≥ 2							
Generous ≥ 3		-0.152					
Generous ≥ 4		0.293					
Generous ≥ 5							
Moral							
Moral ≥ 2		-0.664	-0.124	-0.837		-0.733	-0.510
Moral ≥ 3		-0.334					
Moral ≥ 4							
Moral ≥ 5						0.957	
Easy							
Easy ≥ 2		0.336					
Easy ≥ 3		-0.874					
Easy ≥ 4							
Easy ≥ 5		1.181					
Pressure							
Pressure ≥ 2		0.719					
Pressure ≥ 3							
Pressure ≥ 4							
Pressure ≥ 5		1.738					
Hard to Find							
Hard to Find ≥ 2		-0.421					
Hard to Find ≥ 3		0.478					
Hard to Find ≥ 4		0.080					
Hard to Find ≥ 5		-0.856					
Hard to Read							
Hard to Read ≥ 2		1.525				1.023	0.814
Hard to Read ≥ 3		0.512	0.035				
Hard to Read ≥ 4		-0.105					
Hard to Read ≥ 5		0.041					

Table A.2: Variables and corresponding coefficients resulting from the model selection procedures on the organ donation panel data. Variables are listed along left-hand side, and each column corresponding to one model selection procedure. Standard errors are not included.

	Model	BIC	BIC (restricted)	BIC (objective data)	BIC (objective data, restricted)	Lasso	Lasso (objective data)	MSPE	MSPE (restricted)	MSPE (objective data)	MSPE (objective data, restricted)
1	(Intercept)	0.266	0.266	0.266	0.266	0.372	0.360	0.192	0.297	0.200	0.297
25	Hypothetical	0.695	0.695	0.695	0.695	0.559	0.562	0.525	0.636	0.473	0.636
26	Likelihood										
27	Likelihood ≥ 2										
28	Likelihood ≥ 3					0.013	0.021				
29	Likelihood ≥ 4									0.078	
30	Likelihood ≥ 5										
34	Vicarious							0.244		0.194	
2	Compare										
3	Compare ≥ 2										
4	Compare ≥ 3										
5	Compare ≥ 4									-0.053	
6	Compare ≥ 5										
7	Deserve										
8	Deserve ≥ 2										
9	Deserve ≥ 3							-0.154		-0.184	
10	Deserve ≥ 4							0.183		0.258	
11	Deserve ≥ 5							-0.233	-0.040	-0.144	-0.040
15	Generous										
16	Generous ≥ 2										
17	Generous ≥ 3					0.007	0.009	0.051	0.053	0.069	0.053
18	Generous ≥ 4					0.000	0.000				
19	Generous ≥ 5							0.166			
20	Good										
21	Good ≥ 2										
22	Good ≥ 3										
23	Good ≥ 4										
24	Good ≥ 5										
12	f1										
13	f2										
14	f3									0.002	
31	s1									0.001	
32	s2									-0.001	
33	s3										

Table A.3: Variables and corresponding coefficients resulting from the model selection procedures on the binary game data. Variables are listed along left-hand side, and each column corresponding to one model selection procedure. Standard errors are not included.

How likely do you think the typical person in the United States would be to become an organ donor when filling out this form?

Very unlikely	Somewhat unlikely	Equally likely and unlikely	Somewhat likely	Very likely
○	○	○	○	○

To what extent do you think that the way this form asks about organ donation makes...

	Not at all	Moderately	Somewhat	A fair amount	Very much
an individual feel good about themselves by signing up?	○	○	○	○	○
an individual appear generous to others by signing up?	○	○	○	○	○
an individual feel morally obliged to sign up?	○	○	○	○	○
it easy for an individual to sign up?	○	○	○	○	○
an individual feel pressured to sign up?	○	○	○	○	○

How how hard was it for you to find the organ donation question?

Very easy	Somewhat easy	Neutral	Somewhat hard	Very hard
○	○	○	○	○

How hard to read is the form?

Very easy	Somewhat easy	Neutral	Somewhat hard	Very hard
○	○	○	○	○

Figure B.1: Screenshot of non-choice responses screen in the organ donation application.

In today's session, you will be shown a set of directions and decision problems. Instead of playing the games described, you will be asked to respond with <u>what you think a typical Mechanical Turker would do, think, or feel</u> in each setting.

Your payoffs will be determined <u>not</u> as described on the following page. You will receive the $5.00 participation payment as described in the HIT, plus a $2.00 bonus. Your responses, nor the responses of any other Mechanical Turker, will affect this payment. We are <u>not</u> comparing your answers against the actual behavior of other Turkers to calculate your payoffs; we are simply asking for your opinion.

Nonetheless, please read the directions carefully and take the problems presented seriously.

Figure B.2: Screenshot of page 1 of instructions for non-choice treatments of the binary game. The choice treatment subjects did not see this page.

Remember, you are not playing the game below. You are tasked with thinking about the typical Mechanical Turker in this situation, and responding with what you think they would do.

Overview

In this study, you will make decisions in several different situations (which we will also call "games"). Each decision is independent of each of your other decisions, so that the outcomes in one game will not affect your outcomes in any other game. The each possible outcome of the games results in some amount of money given to or taken from each player.

Description of the Games

The games have two players: Player 1 and Player 2.

There are two types of games that you will play:

- In Type A games, only Player 1 will get to make a decision. So, their decision will completely determine the dollar amounts that the two players receive. Player 1 in this case will choose between two options, which we will label "Left" and "Right". If they choose "Left", they get $u and Player 2 gets $v. If they choose "Right", they get $x and Player to gets $y.

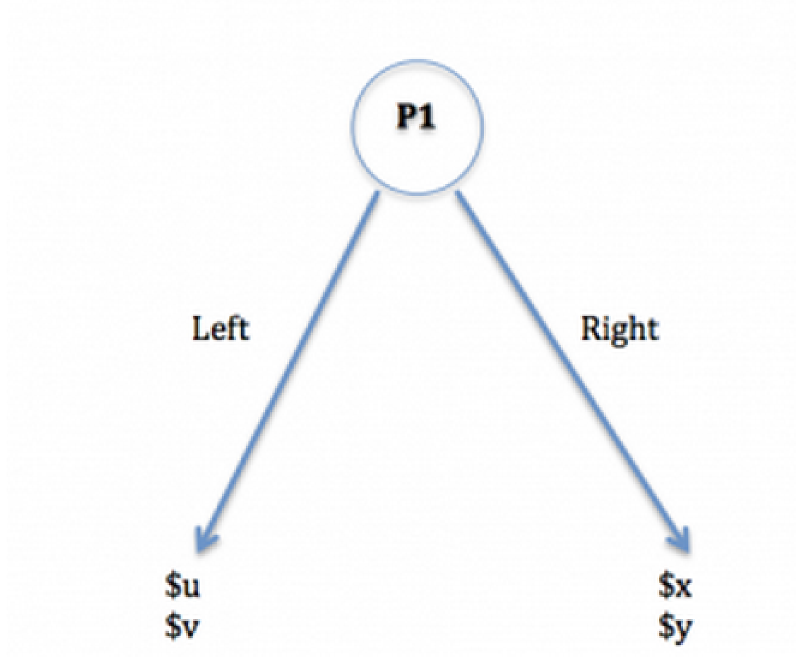
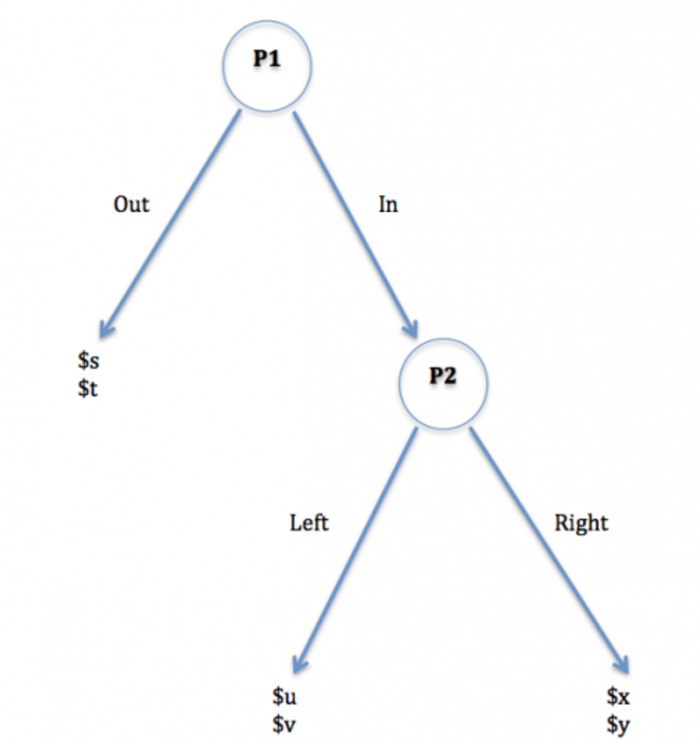


Figure B.3: Screenshot of page 2 of instructions for non-choice treatments of the binary game, part 1. The instructions for the choice treatment were identical, except that the highlighted note at the top of the page was not included.

- In Type B games, Player 1 makes a choice first, between two options that we will label "In" and "Out". If Player 1 chooses "Out", then they get $s and Player 2 gets $t, and game is over. If Player 1 chooses "In", then Player 2 has a chance to respond by choosing "Left" or "Right". If Player 2 chooses "Left", then they get $v and Player 1 gets $u. If Player 2 chooses "Right$, then they get $y and Player 1 gets $x. So, both players have a chance to influence the dollar amounts that they receive. You will play this game both from the perspective of Player 1 and from the perspective of Player 2. When you are making your choices for each game, you will not know what action the other player has chosen.



In summary, you will make three sets of decisions: one set for Type A games, one set for Type B games as Player 1, and one set for type B games as player 2. These set will occur in random order. Within each set, you will play 120 rounds of each game. You will move on to the next game without learning what your opponent has chosen.

Payments
Note that the payoffs you will be shown for each game are in "experimental currency units" (ECU). The conversion ratio is 10 ECU = $1.

Total payments will then be determined as follows:
- Everyone will receive $5.00 for participation, which will be paid automatically upon approval.
- After all participants have submitted, you will be matched into pairs. A computer program will then randomly assign you and your partner to roles (Player 1 and Player 2). Given the responses of you and your partner, final payoffs will then be tallied for one round chosen at random. You will receive any additional payment as a bonus (of between $0.00 and $4.00 depending on your and your partner's choices), to be paid within 5 days.

Figure B.4: Screenshot of page 2 of instructions for non-choice treatments of the binary game, part 2. The instructions for the choice treatment were identical, except that the highlighted note at the top of the page was not included.

Left: 10 for Player 1, 20 for Player 2
Right: 8 for Player 1, 14 for Player 2

Remember, you are Player 1. You must choose either Left or Right.

Choose one option.

- Left: 10 for Player 1, 20 for Player 2
- Right: 8 for Player 1, 14 for Player 2

Figure B.5: Screenshot of a typical decision screen for the choice treatment of the 1-mover game.

Out:
26 for Player 1, 12 for Player 2
In:
If Player 2 chooses Left: 16 for Player 1, 36 for Player 2
If Player 2 chooses Right: 26 for Player 1, 10 for Player 2

Remember, you are Player 1. You must choose either In or Out.

Choose one option.

- Out:
  26 for Player 1, 12 for Player 2
- In:
  If Player 2 chooses Left: 16 for Player 1, 36 for Player 2
  If Player 2 chooses Right: 26 for Player 1, 10 for Player 2

Figure B.6: Screenshot of a typical decision screen for the choice treatment of the 2-mover game from the first player's perspective.

If Player 1 chooses Out:
14 for Player 1, 18 for Player 2
If Player 1 chooses In:
If Player 2 chooses Left: 10 for Player 1, 14 for Player 2
If Player 2 chooses Right: 32 for Player 1, 20 for Player 2

Remember, you are Player 2. You must choose either Left or Right. Your choice will only count in the case that Player 1 chooses In.

Choose one option.

- Left: 10 for Player 1, 14 for Player 2
- Right: 32 for Player 1, 20 for Player 2

Figure B.7: Screenshot of a typical decision screen for the choice treatment of the 2-mover game from the second player's perspective.

Remember, you are not playing the game below. You are tasked with responding with what you hypothetically would do if you were in the situation below.

If Player 1 chooses Out:
34 for Player 1, 6 for Player 2
If Player 1 chooses In:
If Player 2 chooses Left: 22 for Player 1, 10 for Player 2
If Player 2 chooses Right: 30 for Player 1, 24 for Player 2

Remember, you are Player 2. You must choose either Left or Right. Your choice will only count in the case that Player 1 chooses In.

Hypothetically, which option would you choose?

- Left: 22 for Player 1, 10 for Player 2
- Right: 30 for Player 1, 24 for Player 2

Figure B.8: Screenshot of a typical decision screen for the hypothetical question treatment of the 2-player-game from the perspective of the second player. Note the highlighted note at the top of the page. This appeared through the experiment to remind non-choice treatment subjects that they were not making actual choices.

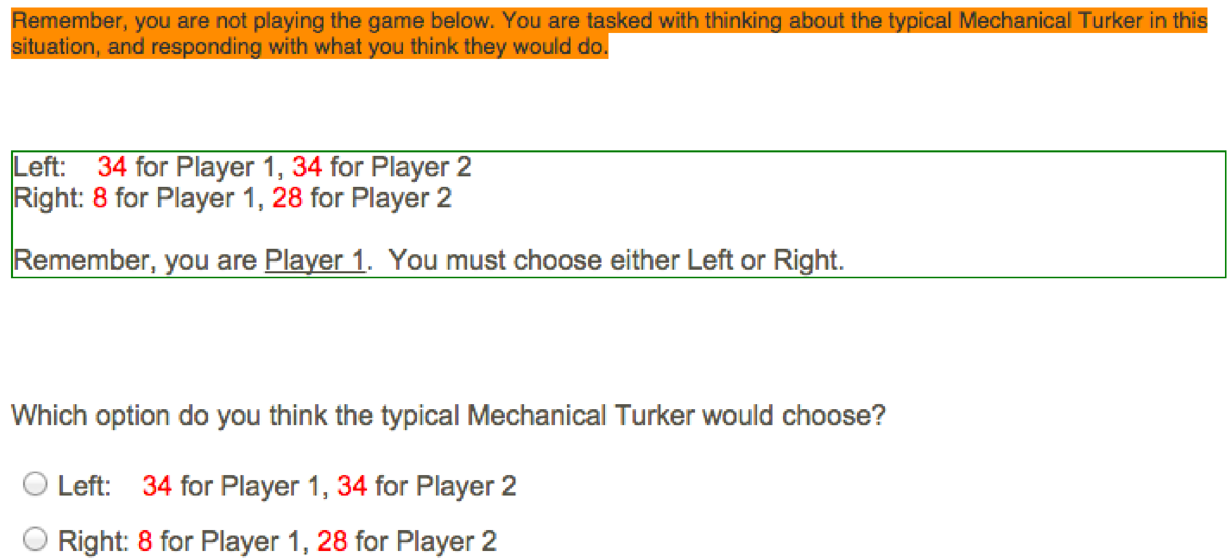


Figure B.9: Screenshot of a typical decision screen for the hypothetical question treatment of the 1-player-game. Note the highlighted note at the top of the page. This appeared through the experiment to remind non-choice treatment subjects that they were not making actual choices.

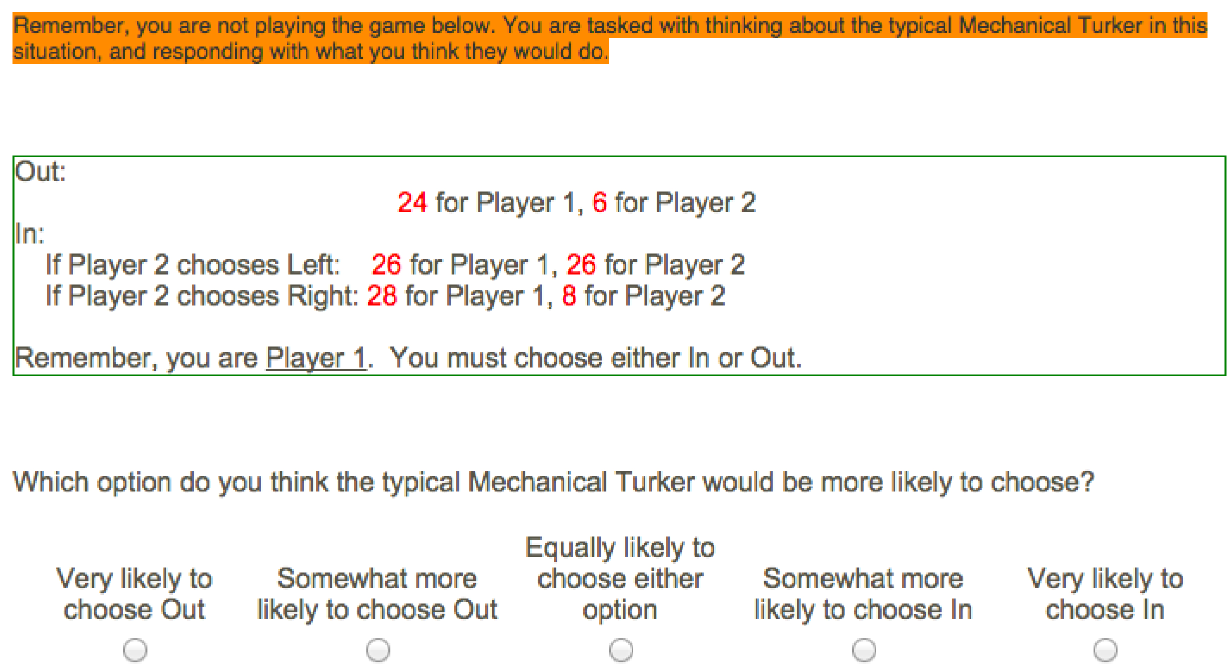
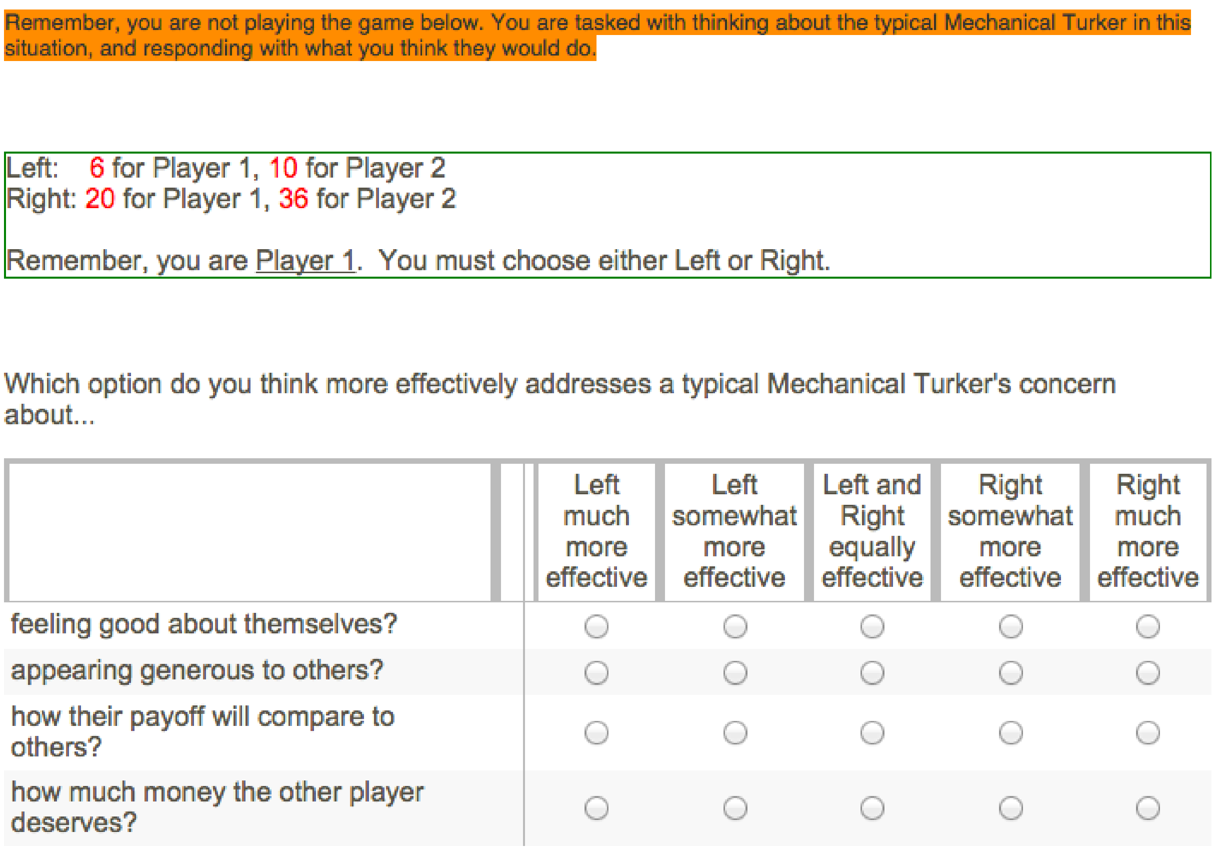


Figure B.10: Screenshot of a typical decision screen for the hypothetical question treatment of the 2-player-game from the perspective of the first player. Note the highlighted note at the top of the page. This appeared through the experiment to remind non-choice treatment subjects that they were not making actual choices.



Remember, you are not playing the game below. You are tasked with thinking about the typical Mechanical Turker in this situation, and responding with what you think they would do.

Left: 6 for Player 1, 10 for Player 2
Right: 20 for Player 1, 36 for Player 2

Remember, you are Player 1. You must choose either Left or Right.

Which option do you think more effectively addresses a typical Mechanical Turker's concern about...

	Left much more effective	Left somewhat more effective	Left and Right equally effective	Right somewhat more effective	Right much more effective
feeling good about themselves?	○	○	○	○	○
appearing generous to others?	○	○	○	○	○
how their payoff will compare to others?	○	○	○	○	○
how much money the other player deserves?	○	○	○	○	○

Figure B.11: Screenshot of a typical decision screen for the subjective questions treatment of the 1-player-game. Note the highlighted note at the top of the page. This appeared through the experiment to remind non-choice treatment subjects that they were not making actual choices.