Reference points and redistributive preferences: Experimental evidence

Jimmy Charité, Raymond Fisman, and Ilyana Kuziemko*

October 23, 2019

Abstract

We explore whether individuals respect others' reference points when acting as social planners. We allow subjects to redistribute unequal, unearned initial endowments between two anonymous recipients. Subjects redistribute twenty percent less when recipients know their initial endowments (and thus may have formed corresponding reference points) than when the recipients do not know their initial endowments, when we observe near-complete redistribution. A second experiment examines whether respect for reference points diminishes support for redistributive taxes: subjects set higher tax rates for someone who became rich last year versus five years ago, indicating respect for more strongly embedded reference levels of income.

JEL Classification Numbers: C9, D63, H21, H23.
Key words: Redistributive preferences, optimal taxation, reference dependence

*We thank Alberto Alesina, Angus Deaton, Stefano DellaVigna, Marc Fleurbaey, Larry Katz, Benjamin Lockwood, David Moss, Howard Rudnick, Stefanie Stantcheva and Matt Weinzierl for helpful discussions, as well as seminar participants at the AEA meetings, Berkeley, Bocconi, UCSB, Chicago, Columbia, Harvard, LSE, NBER, Northwestern, Princeton, UPF, and Stockholm. Financial support from the Tobin Project is gratefully acknowledged. Adith Srinivasamurthy provided invaluable assistance with JavaScript programming and Dana Scott provided excellent research assistance. Charité: Columbia University (email: jc4144@columbia.edu); Fisman: Boston University (email: rfisman@bu.edu); Kuziemko: Princeton University (email: kuziemko@princeton.edu).

1 Introduction

Prospect theory, introduced by Kahneman and Tversky (1979), posits that in many important contexts, individuals behave as though they evaluate options relative to a reference point, typically the status quo (whereas standard utility functions assume that only absolute levels of consumption enter into utility). In particular, individuals often appear to be loss-averse relative to this reference point, in that the reduction in utility from a (relative) loss is greater than the increase in utility from a corresponding gain. The existence of reference-dependent preferences is supported by a large body of research, and has seen wide-ranging application in economists' theoretical and empirical research, on topics ranging from labor supply decisions to trading in financial markets.[1]

If loss aversion captures central features of individual preferences, it is natural to consider whether such concerns are incorporated in redistributive decisions which, almost by definition, create loss (among the rich) and gains (to the poor) across individuals. Voters that consider the loss aversion of their fellow citizens might weigh more heavily the losses (among the rich) than the gain (to the poor), as compared to a voter who did not have such concerns. To the extent that loss aversion does play such a role, it could offer a heretofore unexplored explanation for limited support for redistribution.

The logic is captured in Figure 1. In the textbook set-up of the optimal income tax problem, we begin with two individuals, rich r and poor p, with exogenously determined, unequal initial endowments. If these two individuals have identical, well-behaved utility functions f (with $f' > 0$ and $f'' < 0$, as in panels (a) and (b) of the figure), then a utilitarian social planner would redistribute to the point of full equality, as up until that point the marginal utility benefit to p is greater than the marginal loss to r. In panels (c) and (d), we depict a typical loss-averse utility function. In this case, full redistribution is welfare-reducing, because losses relative to the rich individual's reference point are greater than gains relative to the poor individual's reference point.

We explore the consequences of reference dependence on the demand for redistribution in an experimental setting. Our first experiment tests whether subjects take into account reference points when asked to play the role of social planner, using a design that mirrors the two-person case outlined in the preceding paragraph. Specifically, we confront subjects (referred to in the paper as "redistributors," though such language is never used in the experiment) with a redistributive decision involving two other (randomly selected and anonymous) participants, who received (based on a coin flip) unequal initial endowments, with

[1] See DellaVigna (2009), which notes also that Kahneman and Tversky (1979) is the second-most cited article in all of economics.

one awarded $5 and the other $15. We emphasize that the endowments are random and exogenous, thus removing concerns of deservedness or efficiency, two key factors that could otherwise account for limited redistribution.

To vary redistributors' beliefs about the recipients' reference points, redistributors in the treatment arm were told that the recipients *had already* been informed of their ex-ante allocation; redistributors in the control arm were told that the recipients had *not* been informed of their ex-ante endowments and would thus only be informed of their ex-post allocations. Subjects were then free to redistribute (or not) any whole dollar amount between recipients, subject to the constraint that all ex-post allocations remain non-negative.

Our main finding in the first experiment is that subjects in the reference-point treatment (in which recipients have been informed of their ex-ante endowments) are significantly less likely to reduce inequality between recipients than subjects in the control condition. In our preferred specification, control-group redistributors erase 93 percent of the initial $10 difference between the recipients' endowments, compared to 77 percent in the treatment group.

We calibrate the size of our effect relative to what the literature suggests is one of the key barriers to redistribution: merit, the sense that the well-off have earned (and hence deserve) their income.[2] In one session, we ask respondents to redistribute unequal endowments between recipients whose initial allocations are assigned either by a coin toss (control), or earned by correctly answering SAT questions (treatment). In both cases, the recipients would only know their final endowments, so the coin-toss scenario is identical to the control arm of the main experiment. We find that the control group erases 90% of the initial $10 difference and the treatment group erases 57%. Thus, our hypothesized "reference-dependence" mechanism has an effect size that is half of the luck-versus-merit effect, suggesting that it could be an important and heretofore overlooked explanation for limited redistribution.

There are several candidate explanations for the reference-dependent redistributive decisions of our subjects. Redistributors may (1) view initial allocations as property rights; (2) exhibit a status-quo bias; (3) believe that telling the recipients their initial endowments serves as an implicit promise, and thus reject redistribution on procedural justice grounds; (4) respect subjects' reference points. The near total redistribution in the control condition argues against the most straightforward versions of explanations based on property-rights or status-quo bias. To examine the potential role of procedural justice, we ran a version of the survey which informed subjects that recipients in the treatment condition had been told up front that their initial allocations *could change* before the receipt of payment. This phrasing should minimize procedural justice concerns, as redistributors can now reallocate money

[2]See, e.g., Alesina and Angeletos (2005) and Durante *et al.* (2013).

without worry of breaking an implicit promise to the treatment-group recipients. Our results continue to hold in this variation, indicating that procedural justice is unlikely to be the primary explanation for our treatment effect. We conclude that respect for others' reference points is the theory that best explains all of the results we present.

Having established robust evidence that individuals respect recipients' reference points over small-dollar amounts, we explore via a second experiment whether reference points can help to explain voters' preferences over actual redistributive policies such as income taxation. To the best of our knowledge, no work has formally tested if respondents take into account whether potential taxpayers have grown used to a certain level of income or consumption in assessing redistributive policies.

In this experiment, subjects are randomly assigned to one of two vignettes describing a person whose annual income had increased to $250,000 owing to circumstances unrelated to skill or effort. The only difference between the treatment and control conditions was the length of time that the protagonist had been receiving the higher income—in the treatment condition, he had been earning $250,000 for five years and in the control condition for only one year. If respect for the reference points of the rich deters voters from demanding more redistribution, then the effect should be stronger in situations where that reference point has become more deeply embedded, i.e., for the individual who had already been earning $250,000 for five years.

Consistent with this view, respondents chose a tax rate in the five-year scenario that is roughly 1.7 percentage points lower than in the one-year scenario (for which they chose a rate of 28 percent). As a benchmark, subjects in our data who report having supported Barack Obama in 2012 choose a tax rate just under three percentage points higher than those who report supporting Mitt Romney, suggesting our reference-point effect is over half the Obama-Romney difference and thus comparable in magnitude to other important predictors of tax preferences.

Our paper contributes to a political economy literature on why demand for redistribution falls short of the levels predicted by standard models. In the neoclassical framework, the social planner wants to redistribute because the marginal utilities of the poor are greater than those of the rich, but is deterred from full redistribution by supply elasticities. However, recent work suggests that standard estimates of labor-supply and taxable-income elasticities yield higher marginal tax rates than are typically observed.[3] Moving from the normative to the positive, a democratic political process would similarly be expected to generate a high degree of redistribution: given a right-skewed ex-ante income distribution, the majority of voters have an incentive to support high effective tax rates on the wealthy (Meltzer and

[3]See e.g., Diamond and Saez (2011).

Richard, 1981). Past work aiming to explain this reluctance has focused on the prospect of upward mobility (Benabou and Ok, 2001), the effects of "policy-bundling" redistribution with other, cross-cutting issues (Lee and Roemer, 2006), and the public's misinformation about income inequality (e.g., Ariely and Norton, 2011). Perhaps the clearest challenge to the standard model is individuals' reluctance to fully redistribute exogenous, initial wealth, a seeming contradiction to utilitarianism recently noted by Saez and Stantcheva (2016) and Weinzierl (2014).

We provide a heretofore unexplored explanation for the limited demand for redistribution that, our results suggest, may be quantitatively important. Further, our results offer evidence that the limited-redistribution-of-exogenous-endowments result can be rationalized in a optimal-tax model with a standard utilitarian social-welfare function, albeit with non-standard individual utility functions.

Given that our explanation for limited redistribution builds on insights from social psychology, our paper also contributes to the field of behavioral public finance. Prior studies have have looked at the implications of "behavioral agents" for the taxation of *goods*.[4] Further, some recent theoretical work has calibrated how optimal income tax results may change when agents are susceptible to inattention, present-bias or mental accounting (see Farhi and Gabaix (2015) and Lockwood (2015)).[5] Our paper is quite distinct from existing work, both in its experimental approach and the focus on implications of "behavioral" agents on decisions, such as those over redistribution, that impact *others*.

2 Experimental design

We collect data from eleven distinct sessions (all described in Appendix Table 1). While there were some small differences in wording and presentation among these sessions (which we introduced to test robustness and we describe later in the paper), in each session subjects are asked to allocate a fixed surplus between two other (anonymous) subjects. Five sessions also include our tax survey question (described in Section 5). One session involved a luck-versus-merit rather than the reference-point treatment (described in Section 4).

[4]Past work has examined the implications of addiction, lack of self-control, and inattention. See, e.g., Bernheim and Rangel (2004) , O'Donoghue and Rabin (2006) and Allcott *et al.* (2014), respectively.

[5]There is also a somewhat older literature that examined optimal tax results when agents care about their income *relative to others*. See Boskin and Sheshinski (1978) and Oswald (1983).

2.1 Main redistribution experiment

Each session presents subjects with the opportunity, in most cases hypothetical, to transfer money between two other anonymous participants. In all cases, *the redistributor received only his show-up fee regardless of his decision, so he has no direct self-interested motivation.*

Respondents randomized into the control arm of the survey encountered the following instructions:

> Consider two other participants on MTurk, person A and person B. Based on a coin flip, we have given $5 to person A and $15 to person B.
>
> You can now transfer money between persons A and B. *Persons A and B are not told how much money they were initially given*. If you decide to give Person A $X instead of $5, he or she will simply be told that they have been given $X, and will not know how much they started with. Nor will they know that there is another person (Person B) involved, or that a third party (you) determined the money they received.
>
> Please indicate below what transfer, if any, you would make.

A slider and interactive bar graph (which reflects in real time movements of the slider) appeared directly below these instructions, allowing respondents to easily and transparently transfer money between players. The default position of the slider was on the ex-ante ($5, $15) distribution. Appendix Figure 1 provides a screenshot.

For those randomized into the treatment arm, the second paragraph of the control instructions was modified as follows:

> You can now transfer money between persons A and B. *Persons A and B have already been told how much money we have given them*. If you decide to give Person A $X instead of $5, they will be told that they now have $X instead of $5. They will not know that there is another person (Person B) involved, or that a third party (you) determined the money they received.

Appendix Figure 2 provides a screenshot.

The survey is available for readers to take online.[6]

Two presentational aspects of the main experiment deserve mention. First, the use of the slider requires a default position, which we set to the status-quo allocation of $5 and $15

[6] A version of the online experiment where the redistribution experiment is presented first is available here: `https://az1.qualtrics.com/SE/?SID=SV_b2Tk5a7LuYAk38F&Preview=Survey&BrandID=columbia`.

dollars. As such, we suspect that anchoring bias could lower the amount of redistribution in both treatment and control and thus attenuate any treatment effect. Second, to illustrate clearly the treatment scenario, we write: "If you decide to give Person A $X instead of $5, they will be told that they now have $X instead of $5." By using the poorer person as the illustration, if anything we should prime redistributors to think of the pleasant surprise that the person starting with $5 will experience, again biasing the experiment against finding our hypothesized effect.

The survey ended with standard demographic questions, as well as a question on preferred candidate in the most recent presidential election (the 2012 election for the MTurk subjects, and the 2016 election for the *UAS* subjects). These questions allow us to examine whether our treatment effect is larger for certain groups, and also to compare our sample to more representative populations such as the General Social Survey.

The final questions of the survey relate to whether respondents felt any part of the survey was confusing or biased and also asked for any other feedback they wished to share.

3 Data

Most of our subjects were recruited through MTurk, an online labor market where "requesters" can post *human intelligence tasks* (HITs) to be completed by "workers." Social scientists have increasingly used MTurk to perform experiments and collect survey data (see Kuziemko *et al.*, 2015 and citations therein for a review), and as such we relegate most of the details of our data-collection procedure to the Online Data Appendix. In the Appendix, we describe the steps we take to ensure a subject pool composed of attentive adult Americans (actual humans as opposed to "bots") taking the survey in good faith. As we discuss in the Appendix, beginning in 2018 MTurk saw a surge of bots, but bots were much less of a problem in 2014 when we collect our MTurk data.

While we discuss representativeness and experimental balance later, we note here that the data pass basic reality checks (for example, subjects that report having supported Mitt Romney in 2012 tend to be white and male, mirroring patterns observed in polling data). Over three-quarters of respondents went on to answer an open-ended "feedback" question, with the vast majority providing positive feedback on the survey and writing in colloquial, American English.

In Appendix Table 2 we show how questions on perceived political bias of the survey vary with treatment status. The vast majority of respondents report that the survey felt unbiased, and reports of bias are unrelated to treatment status. Similarly, survey fatigue should not affect our estimates of the treatment effect, as the average number of minutes taken to complete the survey is also independent of treatment status.

While we believe that our MTurk data are of high quality in that real human subjects answered the questions seriously and in good faith, we also replicate our results on the more representative data collected through the *Understanding America Study* (UAS), run by the University of Southern California.

Appendix Table 3 provides details on the subjects who completed our survey experiments, separating between MTurk and *UAS* subjects and comparing them to the nationally representative sample of adults in the General Social Survey. Consistent with past work using MTurk, younger and male subjects are over-represented, and despite being more educated than the representative American adult, subjects have lower households incomes. On these dimensions, the *UAS* data are more representative, quite close to the GSS averages.

Appendix Table 4 provides a larger number of covariates for MTurk and UAS participants (not limited to those that can be compared to the GSS) and reports differences between the control and treatment groups. There is good experimental balance, with no variable showing a statistically significant difference at the five-percent level. In particular, variables we suspect could impact redistributive decisions in our setting—political preferences as captured in presidential election votes—are very similar across experimental arms.

4 Results from the redistribution experiment

Table 1 shows, for the full sample across all nine sessions of the redistribution experiment (including both the MTurk and UAS samples), the main between-subject differences in total redistribution for those first assigned to the control versus those first assigned to the treatment. Recall that redistributing $5 from the "richer" to "poorer" recipient would result in strict equality ($10, $10). Column (1) shows the treatment effect controlling only for session fixed effects. Those in the control group achieve nearly complete redistribution, shifting an average of $4.35 from the richer recipient to the poorer one, or 87% of the redistribution required for strict equality. Recall that the default position of the slider was the status quo ($5 and $15) allocations, suggesting that anchoring or inattention would bias the control group results against inequality-reducing redistribution. Those assigned to the treatment redistribute on average $0.75 (or 17 percent) less than those in the control.

Column (2) drops subjects who finished the survey in less time than one could reasonably be capable of completing it.[7] Column (3) further excludes subjects for whom the $5/$15 experiment was not the first survey item survey (removing subjects that may be contam-

[7]Less than three minutes for the first session, which had fewer follow-up questions, and six minutes for other sessions.

inated by exposure to our income tax survey experiment, discussed in Section 5). In both cases, the control mean and the treatment effect increase slightly. In Appendix Table 5 we show robustness to dropping those who *increase* inequality and to adding covariates to the regression.

We show these results separately by session in Panel (a) of Figure 2. Sessions 2 and 3 are pure replications of Session 1. For Sessions 4 and 5, the \$5/\$15 experiment is identical to that described in Section 2.1, but it appeared *after* our income-tax experiment (described in the next section).

The remaining sessions test the sensitivity of our result to changes in wording, presentation and sampling platform. While in all other cases our subjects dealt with hypothetical situations, in Session 6 they were informed that with ten-percent probability their decisions would be implemented on actual individuals.[8] In another robustness check (Session 7), none of the text was italicized or underlined, and the underlined reminder message (see Appendix screenshots) placed next to the slider was removed.

In Session 8, we altered the language in the treatment condition to convey to redistributors that the initial endowments should not be seen as a promise or obligation to persons A or B. As Leventhal (1980) notes, the procedural justice "rule of adhering to commitments...dictates that fairness is violated unless persons receive that which has been promised to them." To limit redistributors' sense of commitment to an initially promised allocation, we modified the wording of the underlined portion of the instructions to read: "Persons A and B were told how much money they were initially given, though they have also been told that the amount might increase or decrease." All other text was unchanged.

Finally, to ensure that results are not driven by some peculiarity of the MTurk platform, we added questions to the *Undestanding America Survey*, a panel comprised of a representative sample of Americans, and managed by researchers at USC (Session 9).

The stability of estimates in Panel (a) of Figure 2 highlights the robustness of our results to various changes in presentation and different samples. While the estimated treatment effects differ somewhat across sessions, we cannot reject equality of any pair of estimates at even the 10% level.

[8] We cannot know for certain whether our subjects realized that the experiment involved hypothetical individuals, though in the experiment the decision is worded in the conditional ("Please indicate below what transfer, if any, you would make"). Moreover, in the MTurk invitation, subjects were told it was an "opinion" survey, again suggesting hypothetical scenarios. Note that in Session 6, the round involving real stakes, we do not employ deception. We indeed give money to MTurk workers in a manner consistent with the decisions of our subjects. In the case of the treatment condition, recipient MTurk workers were told initially they had \$15 (or \$5) and subsequently told that instead they had whatever amount our subjects allocated them. In the case of the control condition, they are just given sums of money.

Figure 3 shows histograms of the final allocation for the ex-ante "poorer" player, for the treatment and control groups. For both groups, the distribution is bimodal, with the mass at (10, 10) but also a second, shorter peak at (5,15); there is almost no mass between these two points. Thus, most of the treatment effect occurs at the *extensive* margin, i.e., the decision to redistribute at all. The lack of intermediate choices is broadly consistent with subjects responding to recipients' reference points, in that partial redistributions both fail to equalize endowments (as in the standard model) and additionally lead to disappointment for the individual endowed with $15.

Table 1 col. (4) shows these extensive-margin results in a regression framework, in which the dependent variable is an indicator denoting the subject chose no redistribution at all. Just under twelve percent of control-group respondents choose to reallocate zero dollars, compared to just over 25 percent in the treatment group. The remaining columns of Table 1 plus cols. (3) and (4) of Appendix Table 5 show that this result passes the same robustness checks as the main result. Panel (b) of Figure 2 shows that the extensive-margin result is stable across all nine sessions.

To gauge the magnitude of the effect we document above, we compare it to the impact of having endowments earned by merit (score on SAT questions) rather than luck (a coin flip). This comparison is motivated by prior work on redistributive preferences, which has shown that one of the most important determinants of support for redistribution is whether income is seen as resulting from merit versus luck (e.g., Alesina and Angeletos, 2005).

To make this comparison, we run a session in which the control arm was kept the same (the $5 and $15 endowments were determined by a coin flip, and the recipients only know their final allocations), while in the treatment arm respondents were told "the initial amounts given to Persons A and B *were based on their performance on SAT questions* [emph. in original]." As with the control version, the redistributor is told that Persons A and B would only learn of their own final allocation (i.e., we shut down any reference-point effect).

The results are reported in Appendix Table 7. Consistent with an important role of perceived merit in attitudes toward redistribution, the treatment effect is large, indeed larger than in our reference-point experiment. Dividing the coefficients in this table by their analogues in Table 1 suggests that our reference-point effect is about half of the luck-versus-merit difference.[9]

Finally, we examine whether the magnitude of the treatment effect depends on demo-

[9]The SAT treatment may contain an implicit reference point treatment as well: a subject may have a rough sense of their performance on the test, and thus form a reference point over their expected compensation. If redistributors take this expectation into account, we would overestimate the pure "merit" effect and understate the share of the reference-point effect relative to a merit effect.

graphic and background characteristics. Despite the fact that we accumulated a relatively large sample size over our various sessions, of the eleven variables we consider, none of the interaction terms are significant (see Appendix Table 8). This lack of heterogeneity across subjects emphasizes the robustness of the reference-point treatment to individual attributes.

5 Income-tax survey experiment

Our findings thus far suggest that subjects are sensitive to others' reference points in redistributive decisions in laboratory settings over small stakes. To relate our findings more directly to policy-relevant questions, we now turn to results from a survey experiment on preferences over income tax rates.

The question of what constitutes an appropriate income tax rate on well-off households is a much-discussed issue in American politics. A threshold of $250,000 has become a focal point, and surveys often ask about support for higher taxes on households with annual incomes of at least that level.[10] We similarly use this threshold in our experiment's design, described below.

5.1 The survey experiment

Subjects were presented with a vignette describing an individual who had received an unexpected increase in earnings. In most waves, the source of the increase was a corporate takeover of the individual's employer (the "takeover" vignette). Subjects were randomly assigned to either a treatment or control arm, which differed only in the *timing* of when the earnings increase took place.[11]

The "control" vignette took the following form:[12]

> There has been much talk about whether wealthy families are paying their fair share in taxes.
>
> Consider the following person. He has been working for about five years as a regional sales manager at a medium-sized firm. <u>*This year*</u>, his firm was taken

[10]See, e.g., `http://politicalticker.blogs.cnn.com/2012/12/06/trio-of-polls-support-for-raising-taxes-on-wealthy/`.

[11]A version of our survey in which the tax experiment is presented first can be taken at: `https://az1.qualtrics.com/SE/?SID=SV_0MnchCiPWRxAsqV&Preview=Survey&BrandID=columbia`. We do not collect any identifying data—such as IP addresses—via this link.

[12]In the vignettes we reference a tax rate of 22 percent on the "average American," based on NBER Taxsim estimates for combined federal and state income tax, and then add the employee side of payroll taxes.

> over by a larger corporation. While he will be doing the same job as before, to make his pay compatible with the earnings of employees in his position at the larger firm, his salary is now doubled, to $250,000.
>
> If it were up to you, how much of his salary should he pay in taxes? (As a basis of comparison, the average American pays about 22 percent in taxes on the income they make.)

In the treatment variant, we attempt to make the protagonists' reference income of $250,000 more deeply embedded. Instead of receiving the raise just this year, he received it five years ago. Specifically, the second paragraph in the treatment vignette reads:

> Consider the following person. He started five years ago as a regional sales manager at a medium-sized firm. *Soon after starting*, his firm was taken over by a larger corporation. While he did the same job as before, his salary was doubled to make his pay compatible with the earnings of employees in his position at the larger firm. Since then, his annual salary has been roughly steady and is now $250,000.

After reading either the control or treatment version of the vignette, subjects provided their response using a slider positioned immediately below the vignette, with values in the range [0,100] percent and the default set to zero. See Appendix Figures 3 and 4 for screen shots.

In a later session, we changed the reason for the individual's increase in income. In the control version, the second paragraph of the vignette above is replaced with:

> Consider the following person. *This year*, he won the state lottery. As a result, he will receive $250,000 a year for the rest of his life (note that lottery winnings are treated as taxable income).

As before, in the treatment version we simply replaced "*This year*" with "*Five years ago*" and changed the verb tense (from "will receive" to "receives") as appropriate.

Our analysis in this section is motivated by models of habit formation whereby individuals acclimate to conditions—financial or otherwise—over time (see, for example, Bowman *et al.* (1999)). Thus, we conjecture that subjects presented with a vignette where the protagonist has been receiving $250,000 for five years will set a lower tax rate than those with a protagonist that has received high earnings for a short time, and hence is not yet accustomed to it.

5.2 Results

We begin by presenting results based on the pooled sample of both takeover and lottery vignettes. In the first column of Table 2, we present the basic treatment effect, where *Treated* denotes that a subject was presented with the vignette where the protagonist's earnings (via corporate acquisition or lottery) increased five years ago, using the full sample of MTurk participants (even those that did not see the tax survey experiment first). Treated subjects choose a tax rate that is 1.17 percentage points lower than control subjects, significant at the 10 percent level; the control group mean is 28.8 percent. In col. 2, we show results from the subsample where the tax experiment appeared first (so could not be contaminated with the $5/$15 experiment); the treatment effect increases to 1.71 percentage points (with a control group mean of 28.6) and is significant at the five-percent level. In column (3) we include controls, which has little impact on the treatment effect.[13]

The magnitude of this effect is quite large. For example, subjects who supported Obama in the 2012 presidential election choose a tax rate 2.96 percentage points greater than do other respondents (not shown), suggesting that our reference-point effect is over half as large as an "Obama effect."[14]

While we tried to hold everything constant in the treatment and control arms except the strength of the reference point, it is possible that respondents read other differences into the stories. Most likely biases push *against* finding our result. In the five-year (treatment) scenario, the protagonist would have had a greater capacity to accumulate wealth and thus could cover the costs of a greater tax burden more easily than the protagonist who only just received the raise (control scenario). Moreover, we suspect respondents might think it unfair that, purely due to luck, in the five-year (treatment) scenario the protagonist enjoys the large raise after having barely worked for the company, whereas in the one-year (control) scenario he put in his time before getting the raise. Given the greater willingness to redistribute gains due to luck both in our $5/$15 experiment as well as in work cited earlier, these factors should lead respondents to choose a *higher* tax rate for the protagonist in the control scenario.

A concern that pushes in the opposite direction is that individuals confronted with the five-year (treatment) scenario may credit the protagonist with greater merit because he has worked at the larger (higher paying) corporation for longer. While we emphasized that in both cases the individual would receive a raise even though he would be doing the *same job*

[13] A small fraction of subjects choose extreme values: about one percent of subjects selected a tax rate of zero while a few chose rates of 99 and 100 percent. Appendix Table 9 shows the result is robust to different methods of handling outliers.

[14] To estimate this effect, we use the specification from Table 2 (i.e., col. 2) but substitute a "Supported Obama" indicator for the treatment indicator.

as before, in the five-year scenario the individual has apparently managed to fit in at the larger corporation, at least to the point that he has kept his (high-paying) job for five years. In the one-year scenario, the future performance of the protagonist at the new corporation is left unclear.

We thus examine the estimates from the takeover and lottery vignettes separately, as the merit argument only applies in the former case. Col. (4) shows estimates from the takeover vignette and col. (5) the lottery vignette. While the treatment effect for the takeover vignette is larger, both are negative and are statistically indistinguishable from one another (though the lottery effect itself is not distinguishable from zero). Winning the lottery, relative to receiving a raise, is a less likely scenario to consider, and thus may be harder for subjects to evaluate. We thus take it as reassuring that we still find a sizable treatment effect in the hypothesized direction (nearly one percentage point, over one-third of our estimated "Obama" effect), as it suggests that the results from the take-over vignette are unlikely due to subjects attributing merit or desert to those who received the raise five years ago.[15]

While the survey experiments in this section have documented that the strength of an individual's reference point reduces the tax rates assigned by our subjects, the precise mechanism is unclear. Chetty and Szeidl (2007) emphasize consumption commitments (categories of consumption, e.g., owner-occupied housing, that are hard to reduce without incurring large fixed costs) as a reason why utility could drop sharply if consumption falls below the status quo. A literal application of consumption commitments cannot account for the results of the $5/$15 experiment (as the money had not yet been spent and the stakes were modest). However, respondents in the income-tax experiments could well be responding to the perceived consumption commitments of the protagonists in the vignettes. It is plausible that our subjects assume that the person who became rich five years earlier would since have taken on a hefty mortgage and enrolled her children in private schools. This consumption-commitments view presents a possible foundation for the existence of a loss-aversion heuristic, whether for oneself or, as is the case in our experiment, on behalf of others. In this paper we aim to document how asymmetric responses to gains and losses affect redistributive preference more generally and upper-income tax policy in particular, rather than attempting to distinguish among underlying explanations for this behavior.

[15] The preferred *levels* of taxation in the lottery vignette are not higher than in the takeover vignette, somewhat surprising given that respondents typically redistribute less when income is due to merit, and presumably working at a successful company is a more "deserving" source of income than a lottery. Again, our subjects may have viewed the lottery vignette as a rather fanciful scenario.

6 Conclusion

Past work has established that in many contexts, individuals display reference-dependent preferences. We provide robust experimental evidence that individuals who are given the opportunity to redistribute between two recipients with unequal endowments are highly sensitive to the recipients' reference points. When the recipients do *not* know their initial endowments, the redistributor erases nearly the full ex-ante income gap. Redistribution is reduced by nearly twenty percent when the recipients *do* know their ex-ante endowments. This reference-point effect is large, more than half of the effect of having endowments determined via merit rather than luck.

These findings have implications for models of optimal taxation. If losses—even for the wealthy—loom larger than gains, part of the welfare gain from redistribution may be erased. If individuals project their own loss aversion onto others when forming their redistributive preferences, then loss aversion might help explain the gap between voters' stated policy preferences and the more egalitarian normative prescriptions of optimal tax models or the positive predictions from standard political economy models.

References

ALESINA, A. and ANGELETOS, G.-M. (2005). Fairness and redistribution. *American Economic Review*, pp. 960–980.

ALLCOTT, H., MULLAINATHAN, S. and TAUBINSKY, D. (2014). Energy policy with externalities and internalities. *Journal of Public Economics*.

ARIELY, D. and NORTON, M. I. (2011). Building a better americaone wealth quintile at a time. *Perspectives on Psychological Science*, **6** (1), 9–12.

BENABOU, R. and OK, E. (2001). Social mobility and the demand for redistribution: The POUM hypothesis. *Quarterly Journal of Economics*, **116** (2), 447–487.

BERNHEIM, B. D. and RANGEL, A. (2004). Addiction and cue-triggered decision processes. *The American Economic Review*, **94** (5), 1558–1590.

BOSKIN, M. J. and SHESHINSKI, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *The Quarterly Journal of Economics*, pp. 589–601.

BOWMAN, D., MINEHART, D. and RABIN, M. (1999). Loss aversion in a consumption–savings model. *Journal of Economic Behavior & Organization*, **38** (2), 155–178.

CHETTY, R. and SZEIDL, A. (2007). Consumption commitments and risk preferences. *The Quarterly Journal of Economics*, **122** (2), 831–877.

DELLAVIGNA, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, **47** (2), 315–72.

DENNIS, S., GOODSON, B. and PEARSON, C. (2018). Virtual private servers and the limitations of ip-based screening procedures: Lessons from the mturk quality crisis of 2018.

DIAMOND, P. and SAEZ, E. (2011). The case for a progressive tax: from basic research to policy recommendations. *The Journal of Economic Perspectives*, **25** (4), 165–190.

DURANTE, R., PUTTERMAN, L. and VAN DER WEELE, J. J. (2013). Preferences for redistribution and perception of fairness: An experimental study. *Forthcoming, Journal of the European Economic Association*.

FARHI, E. and GABAIX, X. (2015). *Optimal taxation with behavioral agents*. Tech. rep., National Bureau of Economic Research.

KAHNEMAN, D. and TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pp. 263–291.

KENNEDY, R., CLIFFORD, S., BURLEIGH, T., WAGGONER, P. and JEWELL, R. (2018). The shape of and solutions to the mturk quality crisis. *Available at SSRN*.

KUZIEMKO, I., NORTON, M. I., SAEZ, E. and STANTCHEVA, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, **105** (4), 1478–1508.

LEE, W. and ROEMER, J. E. (2006). Racism and redistribution in the United States: A solution to the problem of American exceptionalism. *Journal of Public Economics*, **90** (6), 1027–1052.

LEVENTHAL, G. S. (1980). *What should be done with equity theory?* Springer.

LOCKWOOD, B. (2015). Optimal taxation with present bias.

MELTZER, A. and RICHARD, S. (1981). A rational theory of the size of government. *The Journal of Political Economy*, **89** (5), 914–927.

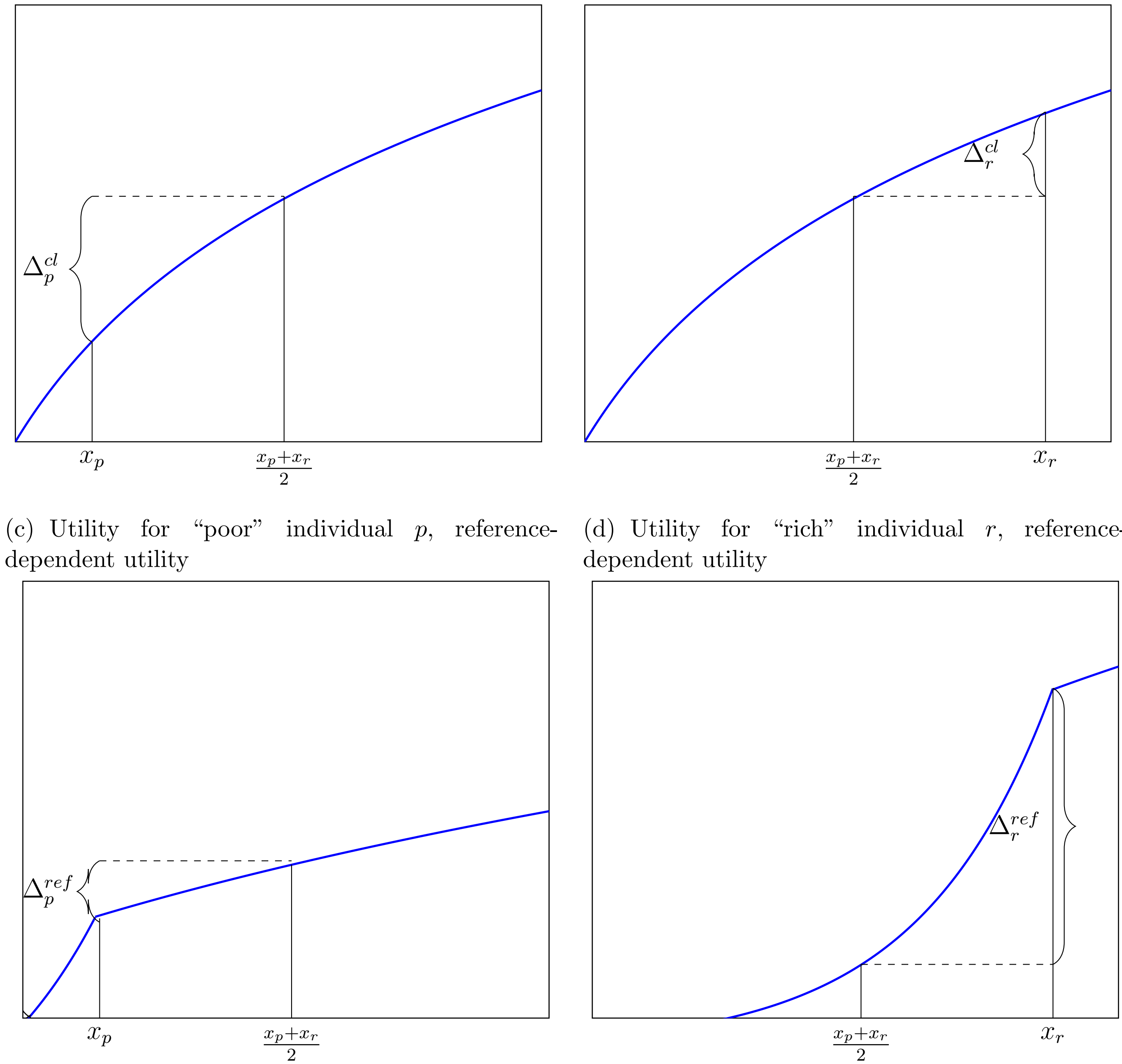
O'DONOGHUE, T. and RABIN, M. (2006). Optimal sin taxes. *Journal of Public Economics*, **90** (10), 1825–1849.

OSWALD, A. J. (1983). Altruism, jealousy and the theory of optimal non-linear taxation. *Journal of Public Economics*, **20** (1), 77–87.

SAEZ, E. and STANTCHEVA, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *The American Economic Review*, **106** (1), 24–45.

WEINZIERL, M. (2014). The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *Journal of Public Economics*, **118**, 128–142.

Figure 1: Redistribution with classical and reference-dependent utility functions



(a) Utility for "poor" individual p, classical utility

(b) Utility for "rich" individual r, classical utility

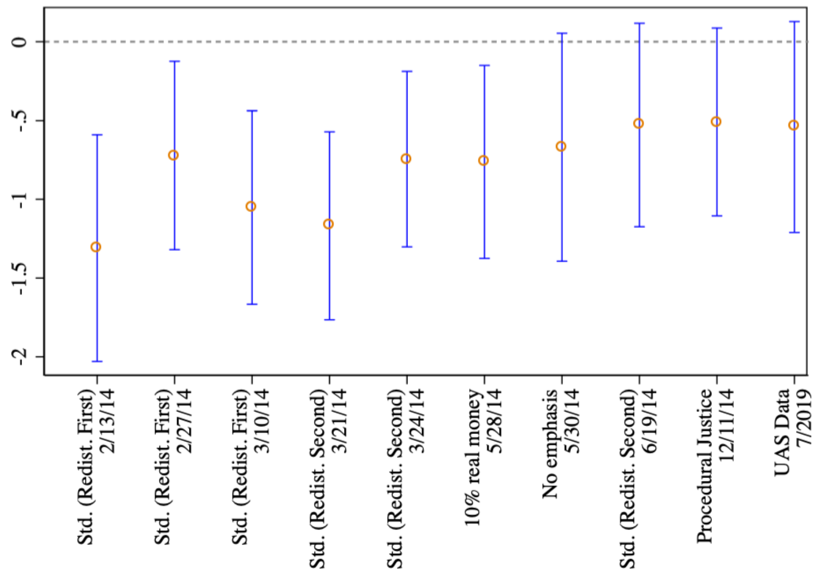
(c) Utility for "poor" individual p, reference-dependent utility

(d) Utility for "rich" individual r, reference-dependent utility

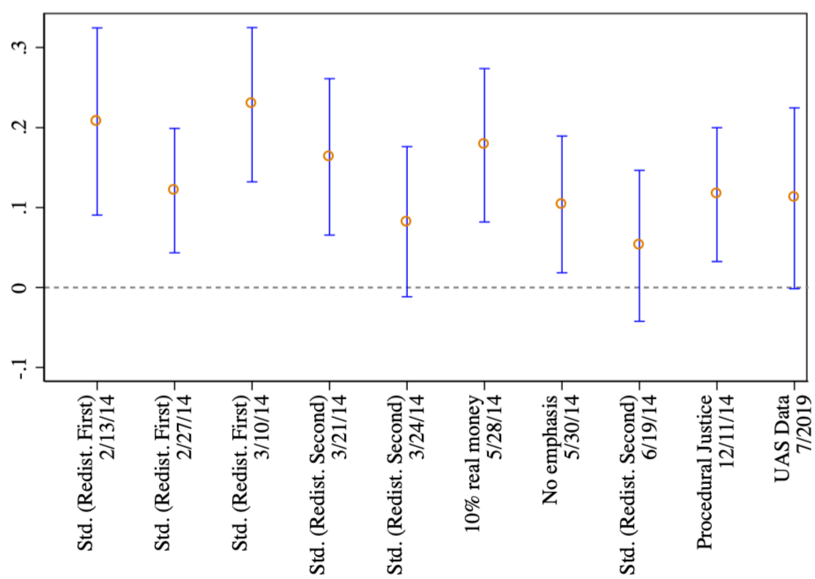
Notes: The figures are drawn by the authors for the sake of illustration. The two top panels show the classic optimal-tax result when initial endowments are exogenous and utility is identical and with standard properties. Under these assumptions, the loss to the rich is smaller than the gain to the poor ($-\Delta_r^{cl} < \Delta_p^{cl}$) and in fact the welfare-maximizing allocation is full equalization at $\frac{x_p+x_r}{2}$. The bottom two panels show how this result can break down under reference-dependent utility, as full redistribution yields lower total utility than the status-quo unequal endowments ($-\Delta_r^{ref} > \Delta_p^{ref}$).

Figure 2: Treatment effects for the reference-point experiment, by session

(a) Total amount redistributed to ex-ante poorer subject

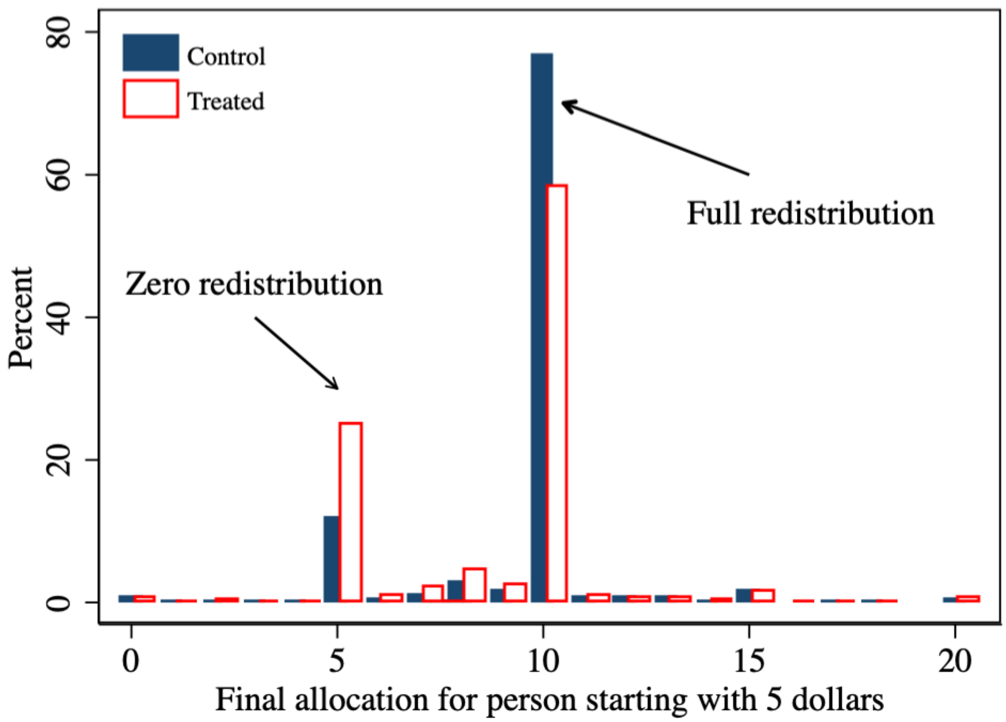


(b) Zero redistributed (dummy variable)



Notes: This figure shows treatment effects and ninety-five-percent confidence intervals, separately for each of the ten sessions of the reference-point experiment (using the sample restrictions in col. 2 of Table 1). We show the treatment effects for two outcomes: (a) the amount redistributed and (b) the (binary) outcome of choosing to redistribute $0. The first five sessions of the experiment have the exact presentation as outlined in Section 2.1. The next four sessions are variants (described in Section 4) included to probe robustness. The final session has the standard presentation, but subjects are taken from the more representative Understanding America Study instead of MTurk. In all sessions, we drop subjects who finish the survey too quickly.

Figure 3: Histogram of ex-post allocations for the ex-ante poorer subject



Notes: Sample used in the figure is that in Column 2, Table 1. Treated refers to subjects who were told the recipients knew their initial allocations and control refers to subjects who were told that the recipients would only know the final allocation.

Table 1: Chosen redistribution in the \$5/\$15 reference-point experiment

	Amount redistributed			Zero redistribution		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated in first stage	-0.747***	-0.789***	-0.829***	0.136***	0.134***	0.158***
	[0.100]	[0.103]	[0.134]	[0.0153]	[0.0158]	[0.0192]
Cont. gp. mean	4.345	4.374	4.682	0.119	0.118	0.0612
Ex. short duration	No	Yes	Yes	No	Yes	Yes
Ex. presented second	No	No	Yes	No	No	Yes
Observations	2352	2194	1227	2352	2194	1227

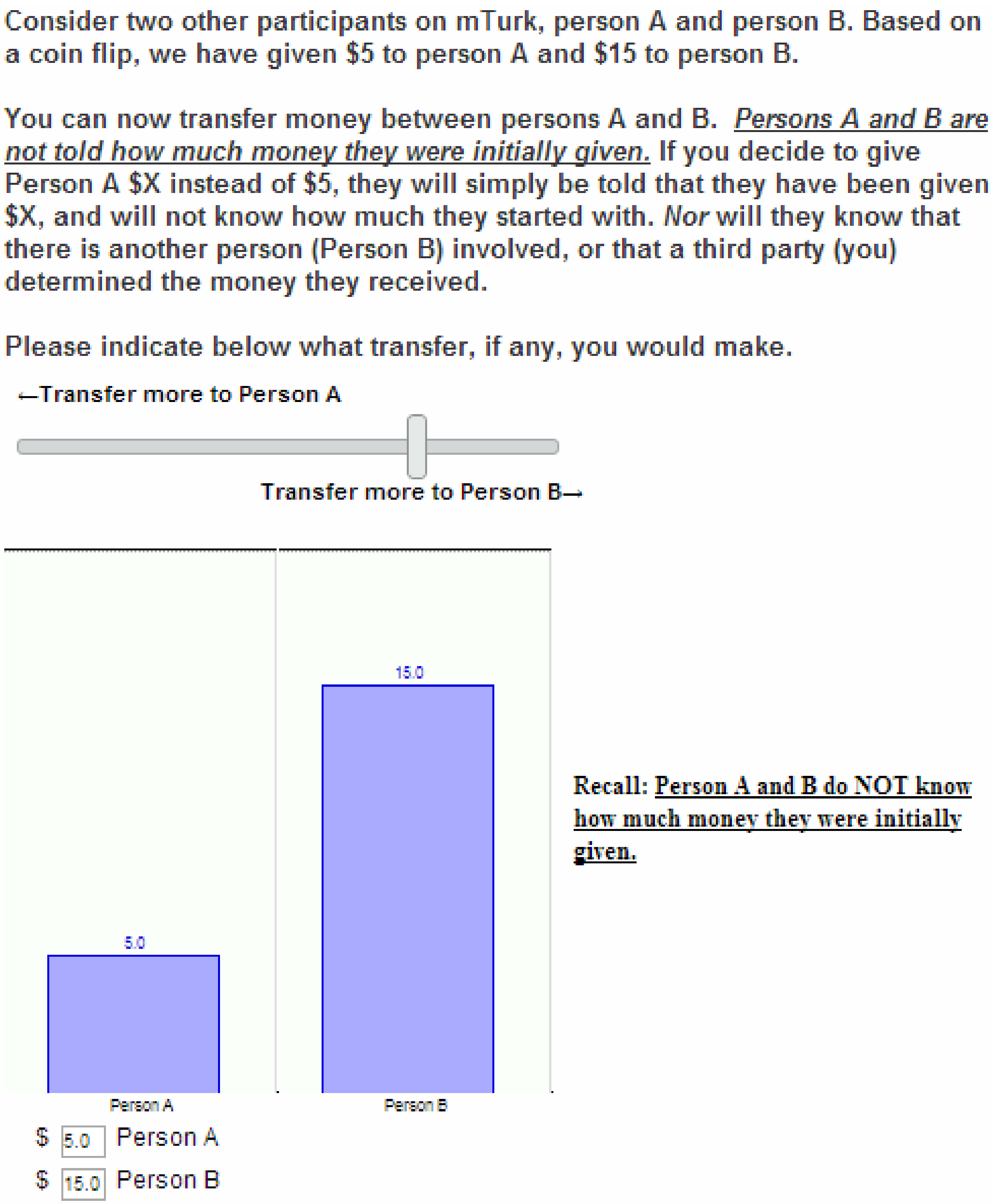
Notes: The outcome in the first three columns is the amount of dollars redistributed from the recipient who starts with \$15 to the recipient with \$5, and the (binary) outcome in the final three columns is whether the subject chose zero redistribution. "Treated" refers to the subject being told that the recipients knew their initial endowments (as opposed to the control group who were told that the recipients would only know their final allocations). All regressions include session fixed effects. Ex. short duration: exclude subjects who finish the survey in a suspiciously short amount of time. Ex. presented second: exclude survey sessions where the main \$5/\$15 reference-point experiment was not presented first. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 2: Preferred average tax rate for person who became rich five versus one year ago

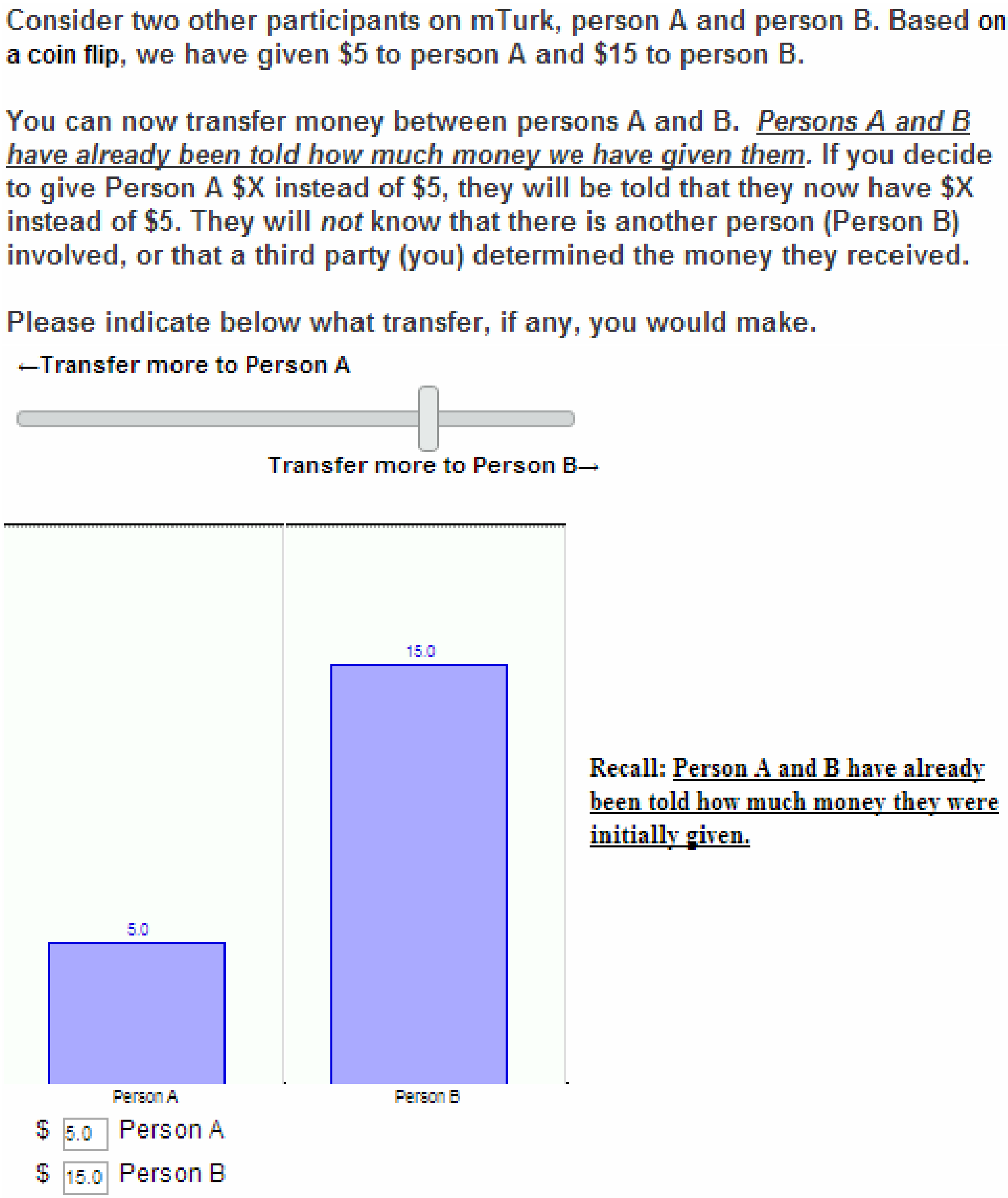
	Dept. var: Chosen tax rate				
	(1)	(2)	(3)	(4)	(5)
Treated (rich for five yrs.)	-0.0117* [0.00644]	-0.0171** [0.00751]	-0.0155** [0.00738]	-0.0197** [0.00868]	-0.00994 [0.0150]
Cont. gp. mean	0.288	0.286	0.284	0.290	0.276
Ex. if presented second	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No
Vignette	Both	Both	Both	Takeover	Lottery
Observations	1097	721	708	532	189

Notes: The outcome variable is the tax rate chosen for the protagonist in the vignette. Treated refers the subject being told that the protagonist's income had increased five years ago (as opposed to the control group, who were told that the protagonist's income increased only one year ago). All regressions include session fixed effects. Subjects who finished the survey very quickly are excluded from the regressions. "Vignette" refers to the brief description of the event that led to the sudden increase in earnings. "Controls" indicates the inclusion of the following covariates: age, female, white, income, student status, full-time status, Obama support, and college degree. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

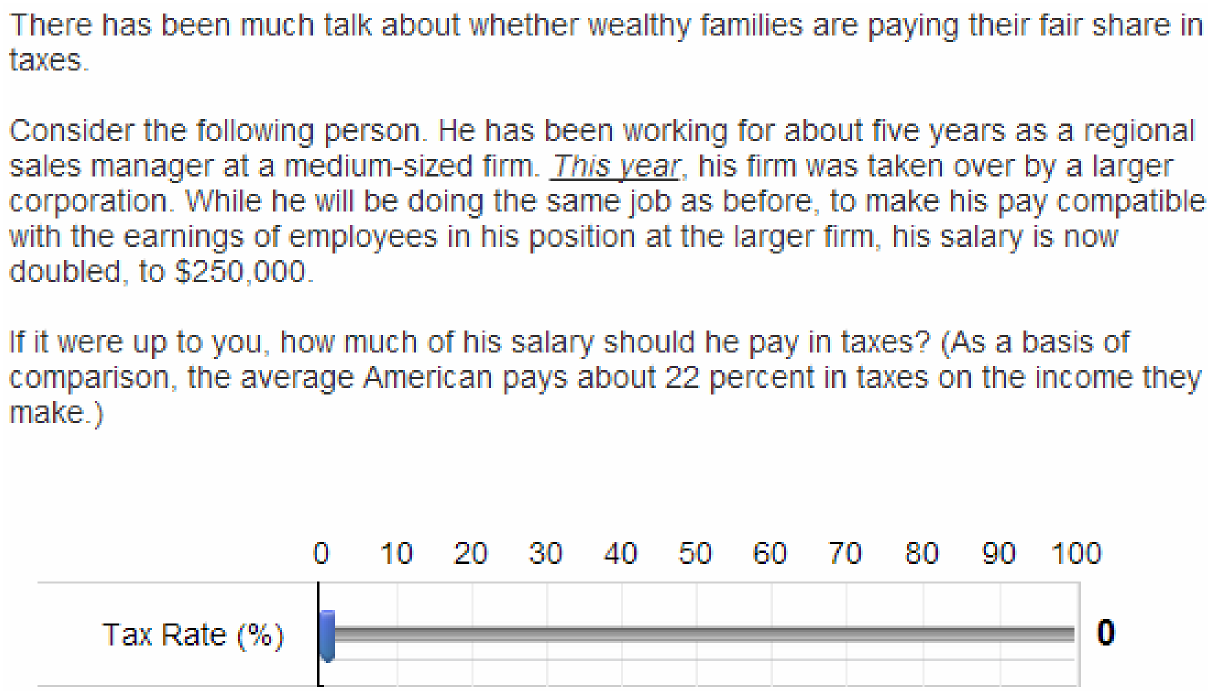
Online Appendix Figure 1: Screenshot of the main redistribution experiment (control arm)



Online Appendix Figure 2: Screenshot of the main redistribution experiment (treatment arm)



Online Appendix Figure 3: Screenshot of income-tax experiment (control arm)

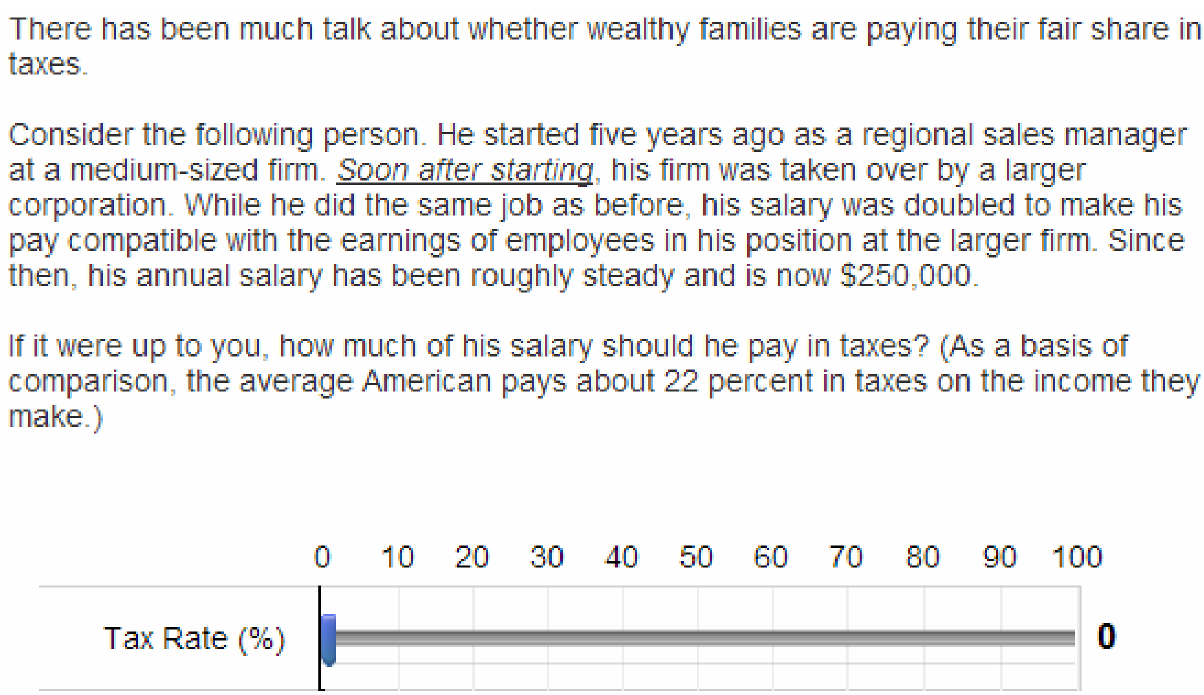


There has been much talk about whether wealthy families are paying their fair share in taxes.

Consider the following person. He has been working for about five years as a regional sales manager at a medium-sized firm. *This year*, his firm was taken over by a larger corporation. While he will be doing the same job as before, to make his pay compatible with the earnings of employees in his position at the larger firm, his salary is now doubled, to $250,000.

If it were up to you, how much of his salary should he pay in taxes? (As a basis of comparison, the average American pays about 22 percent in taxes on the income they make.)

Online Appendix Figure 4: Screenshot of income-tax experiment (treatment arm)

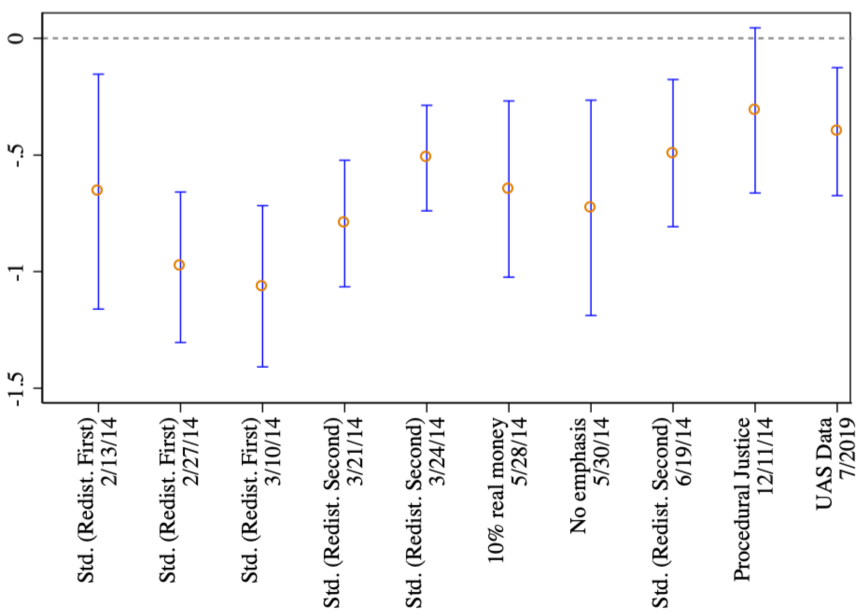


There has been much talk about whether wealthy families are paying their fair share in taxes.

Consider the following person. He started five years ago as a regional sales manager at a medium-sized firm. *Soon after starting*, his firm was taken over by a larger corporation. While he did the same job as before, his salary was doubled to make his pay compatible with the earnings of employees in his position at the larger firm. Since then, his annual salary has been roughly steady and is now $250,000.

If it were up to you, how much of his salary should he pay in taxes? (As a basis of comparison, the average American pays about 22 percent in taxes on the income they make.)

Online Appendix Figure 5: Within-subject treatment effects for the $5/$15 reference-point experiment



Notes: This figure replicates the analysis from Figure 2 panel (a), but uses within-subject variation instead of between-subject variation in treatment status. Most of the analysis in the paper uses as identification that subjects were randomly assigned the treatment or control version of the $5/$15 experiment (between-subject variation). But we also run the "reserve experiment" by then presenting the control group the treatment version (and vice versa), giving us within-subject variation.

Online Appendix Table 1: Survey Session Details

Date	Observations Total	Observations Unique	First Exper.	Version of $5/$15 exp.	Income tax exp.?	Platform
Reference point experiments						
Feb 13, 2014	187	187	$5/$15	Standard	No	MTurk
Feb 27, 2014	312	295	$5/$15	Standard	No	MTurk
Mar 10, 2014	301	250	$5/$15	Standard	No	MTurk
Mar 21, 2014	352	282	Tax	Standard	Yes	MTurk
Mar 24, 2014	374	303	Tax	Standard	Yes	MTurk
May 28, 2014	312	207	$5/$15	Real Money	Yes	MTurk
May 30, 2014	332	216	$5/$15	No Emphasis	Yes	MTurk
Jun 19, 2014	314	200	Tax	Standard	Yes	MTurk
Dec 11, 2014	307	196	$5/$15	No Promises	No	MTurk
June–July 2019	308	308	n.a.	Standard	No	UAS
Luck v. merit experiment						
Apr 25, 2014	321	228	$5/$15	Luck/Merit	No	MTurk

Notes: Total observations and analysis sample observations differ because in almost all analyses we drop anyone who took a previous survey. We collect data on two different platforms: MTurk (Amazon Mechanical Turk) and the UAS (Understanding America Study) panel. On MTurk, each session ran for one or two hours. On UAS, the data were collected from June 24 to July 19, 2019. Further details on the wording used in each session can be found in the text (Sections 2.1 and 4).

Online Appendix Table 2: Assessing perceptions of bias and survey fatigue, by redistribution experiment survey arm

	(1) LW bias	(2) RW bias	(3) No bias	(4) Minutes
Treated in first stage	0.00943 [0.0143]	-0.000315 [0.00933]	-0.0132 [0.0168]	0.148 [0.307]
Cont. gp. mean	0.103	0.0423	0.850	12.26
Observations	1883	1883	1883	1733

Notes: The outcome variables in the first two columns is a binary variable for the subject reporting they felt that the survey experiment exhibited left-wing or right-wing bias, respectively. Subjects who finished the survey very quickly were not included in these regressions. Also, we did not ask the bias questions to the UAS sample and thus they are also excluded.

Online Appendix Table 3: Basic summary statistics in MTurk and UAS sample compared to the General Social Survey

	MTurk	UAS	GSS
Age	33.03	53.23	47.44
Female	0.44	0.57	0.55
White	0.77	0.87	0.73
Black	0.07	0.08	0.15
Hispanic	0.06	0.08	0.16
Asian	0.08	0.03	
Has at least college education	0.45	0.45	0.32
Subject Income (Thousands of Dollars)	49.53	72.96	71.56
Supported Obama in 2012	0.64	.	0.62
Supported Clinton in 2016	.	0.48	0.55
Observations	2029	315	7753

Notes: Col. 1 includes all ten sessions of the reference-point experiment conducted in MTurk. Col. 2 includes the session conducted on the UAS sample. Col. 3 includes all adults in the 2014, 2016, and 2018 GSS (weighted with the provided individual-level weights), except that only the 2014 and 2016 surveys are used for the Obama question and only the 2018 survey is used for the Clinton question. "Income" refers to household income (in units of $1,000).

Online Appendix Table 4: Further summary statistics and experimental balance

	Means				Observations	
	Treat	Control	Diff.	p-val	Treat	Control
Age	35.96	35.71	0.252	0.659	1189	1127
Female	0.460	0.457	0.00309	0.882	1189	1127
White	0.792	0.783	0.00965	0.570	1189	1127
Black	0.0706	0.0781	-0.00744	0.495	1189	1127
Hisp	0.0563	0.0603	-0.00399	0.682	1189	1127
Asian	0.0732	0.0665	0.00662	0.532	1189	1127
Income (in thousands of dollars)	52.74	52.01	0.738	0.669	1189	1126
Full-Time Work	0.442	0.430	0.0120	0.559	1189	1127
Part-Time Work	0.143	0.123	0.0196	0.165	1189	1127
College Degree	0.451	0.445	0.00537	0.795	1189	1127
Supported Dem. in last pres. election	0.628	0.620	0.00729	0.724	1144	1072

Notes: The first column displays means for those randomized into the treatment version of the \$5/\$15 money-transfer experiment (where recipients do know their original endowment) and the second column displays the means for the control version (where recipients do not know their original endowment). The third column displays the difference and the fourth column the p-value associated with $H_0 : Diff = 0$ for the difference between treatment and control groups. The final two columns give sample sizes for the treatment and control groups. This table includes only the ten sessions where the experiment includes the \$5/\$15 reference-point experiment. It does not include the \$5/\$15 luck-merit experiment (analyzed in Appendix Table 7). The experimental balance table for this experiment is available upon request. The number of observations differ across variables because of subject non-response.

Online Appendix Table 5: Robustness of main results from reference-point experiments

	Amount redistributed			Zero Redistribution		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated in first stage	-0.789*** [0.103]	-0.778*** [0.0967]	-0.766*** [0.104]	0.134*** [0.0158]	0.136*** [0.0160]	0.127*** [0.0158]
Cont. gp. mean	4.374	4.470	4.421	0.118	0.120	0.113
Controls	No	No	Yes	No	No	Yes
Ex. incr. ineq	No	Yes	No	No	Yes	No
Observations	2194	2167	2088	2194	2167	2088

Notes: This table presents additional specifications of the main results in Table 1. All regressions include session fixed effects. Controls: age, female, white, black, Hispanic, asian, income, student status, full-time status, part-time status, Democratic-candidate support in the most recent presidential election, and college degree. Ex. short duration: exclude subjects who finish the survey in a suspiciously short amount of time. Ex. presented second: exclude survey sessions where the main redistribution experiment was not presented first. Ex. incr. ineq: exclude subjects who choose to make inequality-*increasing* reallocations. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Online Appendix Table 6: Within-subject results

	Dept. var: Amount redistributed		
	(1)	(2)	(3)
Treatment stage	-0.661***	-0.600***	-0.715***
	[0.0538]	[0.0825]	[0.0706]
Dept. var. mean	3.965	3.909	4.016
Sample	All	T $\rightarrow C$	C $\rightarrow T$
Observations	4306	2044	2262

Notes: All regressions include respondent fixed effects so as to make use of within-subject variation (and thus the number of observations are twice the number in the baseline, between-subject specifications). Subjects who finished the survey very quickly were excluded. $C \rightarrow T$ denotes the subsample that was first randomized into the control scenario and *then* the treatment scenario. $T \rightarrow C$ denotes the subsample that was first randomized into the treatment scenario and *then* the control scenario. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Online Appendix Table 7: Luck (control) versus merit (treatment), between-subject results

	Amount redistributed		Zero redistribution	
	(1)	(2)	(3)	(4)
Treated in first stage	-1.806***	-1.661***	0.194***	0.178***
	[0.360]	[0.366]	[0.0576]	[0.0590]
Cont. gp. mean	4.515	4.510	0.136	0.137
Controls	No	Yes	No	Yes
Observations	206	205	206	205

Notes: This table presents results parallel to cols. (2) and (5) of Table 1 except the data are taken from the "luck-versus-merit" session of the $5/$15 experiment (see text for details). "Treated" in this experiment means that the subject was told that the hypothetical individuals had earned their endowment by answering SAT questions (versus those in the control who were told they were allocated based on a coin toss). "Controls" indicates the inclusion of the following covariates: age, female, white, income, student status, full-time status, Obama support, and college degree.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Online Appendix Table 8: Differential treatment effects from between-subject results

	Dept. var: Amount redistributed											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated in	-1.377***	-0.958***	-0.925***	-0.894***	-0.931***	-0.862***	-0.931***	-0.877***	-0.926***	-0.921***	-0.834***	-0.784***
first stage	[0.321]	[0.138]	[0.214]	[0.106]	[0.105]	[0.106]	[0.166]	[0.108]	[0.135]	[0.110]	[0.137]	[0.169]
Tr. x Age	0.0142											
	[0.00910]											
Tr. x		0.124										
Female		[0.205]										
Tr. x			0.0291									
White			[0.243]									
Tr. x				-0.106								
Black				[0.386]								
Tr. x Hisp					0.498							
					[0.438]							
Tr. x						-0.546						
Asian						[0.390]						
Tr. x							0.000000592					
Income							[0.00000267]					
Tr. x								-0.230				
Student								[0.330]				
Tr. x									0.0558			
Fulltime									[0.208]			
Tr. x										0.133		
Partime										[0.294]		
Tr. x											-0.153	
College											[0.207]	
Tr. x												-0.186
Supported Democrat												[0.213]
Cont. gp. mean	4.573	4.573	4.573	4.573	4.573	4.573	4.573	4.573	4.573	4.573	4.573	4.573
Observations	2093	2093	2093	2093	2093	2093	2093	2093	2093	2093	2093	2093

Notes: All regressions include session fixed effects and the following covariates: age, female, white, income, student status, full-time status, Obama support, and college degree. Subjects who finished the survey very quickly and were not presented the distribution experiment first were excluded from these regressions. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Online Appendix Table 9: Replicating the main tax result with various outlier adjustments

	Dept. var: Chosen tax rate				
	(1)	(2)	(3)	(4)	(5)
Treated (rich for five yrs.)	-0.0168***	-0.0300***	-0.0169**	-0.0158**	-0.0127**
	[0.00630]	[0.00883]	[0.00671]	[0.00739]	[0.00562]
Cont. gp. mean	0.288	0.286	0.286	0.287	0.306
Outlier adjustment	Drop	Qreg	Winsorize	Drop 0s	Drop regr.
Observations	694	721	721	717	602

Notes: This table subjects the results in Table 2 to various outlier adjustments. In col. 1, outliers below the bottom fifth percent and above the 99th percentile are dropped (the asymmetry is due to a small mass point at zero, the default position of the slider with which subjects indicated their preferred tax rate). "Qreg" refers to median regression on the entire sample. "Winsoring" winsorizes the outliers in Col. (1) instead of dropping them. Col. (4) merely drops those who choose a zero tax rate while col. (5) drops anyone who chooses a regressive tax rate (i.e., a rate less than the average rate of 22 percent). $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Data Appendix

Much of the data in this paper is collected via Amazon Mechanical Turk (MTurk), and in this short data appendix we detail our data collection process. MTurk allows "requesters" to find "workers" to complete "human intelligence tasks" (HITs). As of the summer of 2014 (when much of our data were collected), MTurk advertised that requesters could "access more than 500,000 workers."

We registered as a requester and posted the following HIT: "Short (less than ten minutes) opinion survey on a variety of topics." We tried to use a neutral description that would limit selection bias while also giving workers an honest description of the task. Compensation was set to $1, approximately minimum wage assuming that subjects took seven minutes to complete it. Actual median completion time was 10.1 minutes, implying an hourly wage of $6.09. Though we cannot find official data on average wages on MTurk, worker forum threads suggested that, at the time, we were paying a generous wage (and when we posted requests for 300 survey takers, the sample was generally gathered within an hour).

Each worker logs in with an MTurk ID. We collected MTurk data over eleven separate sessions, dropping workers who had taken a previous survey with the same ID, to ensure a fresh sample each time (though our main results hold when we keep repeat-takers in the sample, see Appendix Table 10).[16]

An important concern with MTurk is the possibility of "bots," algorithms that masquerade as humans. We thus begin each survey with a "captcha" (non-standard writing difficult for computers to interpret). Furthermore, when we ask respondents for feedback at the end of the survey, essentially the only negative comment was that the captchas were "hard," suggesting that algorithms would have a difficult time parsing them. Nonetheless, to address worries that robots learn how to read "captchas," we hand-sketched pictures of a cat, dog, horse and panda bear, and respondents in later rounds were asked to answer multiple-choice questions of the form: "this is a picture of...." after seeing these sketches. If robots still remain in our sample after these checks, they would attenuate any treatment effect. Importantly, all of our MTurk data are collected well before 2018, when many researchers found that MTurk became infested with "bots."[17]

To limit heterogeneity of the sample, we collect all data on workdays during daylight hours on the U.S. East Coast. Individuals were automatically prompted for a response when they tried to skip questions (to discourage robots or inattentive respondents). Given our focus in parts of the survey on American tax policy, we limited the survey's availability to those with U.S. billing addresses, and asked respondents to confirm their U.S. residency. To further ensure the attentiveness of our subjects, we limit respondents to those with positive ratings from at least ninety percent of past requesters.

[16]If workers have multiple IDs then some individuals may participate in multiple sessions. Outside of surveys (which appear to make up a very small share of all HITs), there would seem to be little financial incentive for MTurk workers to create multiple MTurk IDs, but we cannot completely eliminate the possibility that some have done so.

[17]See, e.g., Kennedy *et al.* (2018) and Dennis *et al.* (2018). These researchers refer to the summer of 2018 as the "MTurk quality crisis."

Online Appendix Table 10: Main between-subject redistribution and tax results, includes subjects that participated in multiple batches

	Amount redistributed	Chosen tax rate
	(1)	(2)
Treated	-0.785***	-0.0138**
	[0.119]	[0.00564]
Cont. gp. mean	4.595	0.289
Observations	1577	908

Notes: The first column replicates the analysis from col. 2 of Table 1 and the second replicates the analysis from col. 2 of Table 2. “Treated” in col. 1 refers to the $5/$15 experiment (i.e., deciding how much to redistribute when the recipients know their initial endowments as opposed to when they do not, as in control condition) and “Treated” in col. 2 refers to the income tax survey experiment (i.e., deciding on ideal tax rate for person who received income increase five years ago as opposed to one year ago, as in control condition). $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$