Reference points and redistributive preferences: Experimental evidence

Jimmy Charité, Raymond Fisman, and Ilyana Kuziemko*

July 31, 2016

Abstract

If individuals evaluate outcomes relative to the status quo, then a social planner may limit redistribution from rich to poor even in the absence of moral hazard. We present two experiments suggesting that individuals, placed in the position of a social planner, respect the reference points of others. First, subjects are given the opportunity to redistribute unequal, unearned initial endowments between two anonymous recipients. They redistribute significantly less if the recipients know the initial endowments (and thus may have formed corresponding reference points) than if the recipients do not know their endowments, when we observe near-complete redistribution. In a separate experiment, respondents are asked to choose a tax rate for someone who (due to luck) became rich either five years or one year ago. Subjects faced with the five-year scenario choose a lower tax rate, indicating respect for the more deeply embedded (five-year) reference point. Our results thus suggest that respect for reference points of the wealthy may help explain why voters demand less redistribution than standard models predict.

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. Charité: Columbia University (email: jc4144@columbia.edu); Fisman: Boston University (email: rfisman@bu.edu); Kuziemko: Princeton University (email: kuziemko@princeton.edu).

1 Introduction

In a standard welfarist model of taxation with no moral hazard, a utilitarian social planner will aspire to equalize wealth across all members of society (Mirrlees, 1971). While moral hazard makes the social planner redistribute less, recent work suggests that standard estimates of labor supply elasticity yield higher marginal tax rates than are typically observed.[1] 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 Richard, 1981).

However, many researchers have observed that neither survey-based expressions of policy preferences nor actual policies consistently reflect the more egalitarian predictions of standard models (see, e.g., Roemer, 1998 and citations therein). Perhaps the clearest challenge to the standard model is individuals' reluctance to fully redistribute exogenous, initial wealth, a contradiction to utilitarianism recently noted by Saez and Stantcheva (2016) and Weinzierl (2014).

Scholars have proposed a number of explanations for the limited demand for redistribution. Past work has focused on the prospect of upward mobility (Benabou and Ok, 2001), the effects of "policy-bundling" redistribution with other, cross-cutting issues (typically race in the U.S. context, see Lee and Roemer, 2006), and the public's misinformation about income inequality (see Ariely and Norton, 2011 on the level of misinformation, though see also Kuziemko *et al.* (2015) on the limited effect of information on policy preferences).[2]

This paper proposes an additional, heretofore unexplored (to the best of our knowledge) explanation for the limited demand for redistribution, based on reference-dependent preferences. Whereas standard utility functions assume that only absolute levels of consumption enter into utility, decades of research support the view that in many important contexts individuals behave as though they evaluate options *relative to a reference point*, typically the status quo. 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. Further, to the extent that voters recognize the reference-dependence of others (and incorporate others' welfare into their own utility),

[1]See Diamond and Saez (2011), who argue that a utilitarian social welfare function would yield top marginal tax rates over 70 percent given consensus estimates of labor supply elasticity.

[2]There is a related literature (see Harms and Zink, 2003 for a review) that examines why policy *outcomes* may be more regressive than voters' preferences even in a democracy (e.g., due to efforts by lobbying groups), but here we focus on the question of why voters' *preferences* might be more regressive than the predictions of standard models suggest they would be.

they may demand limited redistribution from rich to poor, even with exogenous, unearned initial endowments.

We explore the consequences of reference dependence on the demand for redistribution in an experimental setting where endowments are random and exogenous, thus removing concerns of deservedness or efficiency, two key factors that could limit redistribution. Our subjects (referred to for convenience in the paper as "redistributors," though such language is never used in the experiments) were recruited via Amazon's Mechanical Turk (mTurk), a rapidly growing online labor market, which we describe in more detail later in the paper. The experiment confronts subjects with a redistributive decision involving two other (randomly selected and anonymous) mTurk participants, who received (based on a coin flip) initial endowments, with one assigned to receive 5 dollars and the other 15. For most redistributors, the decision was presented as hypothetical; for a subset we confirmed that the results hold when subjects are informed that with 10 percent probability their decisions will be implemented for real stakes. In all cases, redistributors are paid only their show-up fee and thus have no personal financial stake in their decision.

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. Each redistributor was presented with both treatment and control conditions, with the ordering chosen at random. For most of our analysis we use between-subject estimates of the reference dependence treatment based on subjects' responses to the first condition they are presented with.

Our main finding is that subjects in the reference-point treatment (where subjects are told that 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 94 percent of the initial 10 dollar difference between the recipients' endowments, while only 77 percent of endowment differences are removed through reallocation in the treatment group. Our estimates hold across a large number of robustness checks—dropping those who finish the survey in a suspiciously short amount of time, dropping those who choose to increase inequality, changing presentational aspects of the experiment, and moving from hypothetical to "real stakes" scenarios. While we prefer between-subject comparisons, which are based on the version the respondent sees first, the within-person estimates are very similar and also highly significant. Further, the estimates remained stable as we collected data on these variants over the course of nearly a year.

It bears repeating that we do not view our reference-point mechanism as supplanting other explanations for limited redistribution. We take care to calibrate the size of our effect relative to what the literature suggests is one of the key barriers to redistribution: the sense that the well-off have earned (and hence deserve) their income.[3] In one session, we asked respondents to redistribute unequal endowments between recipients whose initial allocations were assigned either by a coin toss, or earned by correctly answering Scholastic Assessment Test (SAT) questions, the main college admissions exam in the US. 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. Consistent with past work, we find strong effects of deservedness: in the SAT treatment, redistributors close 56% of the ex-ante gap. Our hypothesized "reference-dependence" mechanism thus has an effect size that is over 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. We present evidence against each of first three mechanisms and conclude that our results are most consistent with our redistributors respecting the reference points of the recipients.

While the main hypothesis we seek to test is that individuals acting as social planners redistribute less when recipients have formed reference points, we also hypothesized that this effect would be largest among social planners who are *themselves* more loss averse. A key challenge in testing this hypothesis is the lack of any consensus on how to measure an individual's intensity of loss aversion. The vast majority of studies on loss aversion use *between-subject* designs to establish the *existence* of loss aversion. We need individual-level measures in the *intensity* of loss aversion to test whether this individual trait predicts differential treatment effects. We propose several such measures, each with its own shortcomings, which we discuss in greater detail in Section 3.2. Overall, we find at best mixed evidence that a subject's own loss aversion drives the treatment effects that we document. This 'non-result' may reflect a lack of any true effect, or instead the difficulty in generating an appropriate measure individual-level intensity of loss aversion (in fact, our proposed measures of loss aversion are not highly correlated within person).

[3]See, for example, Alesina and Angeletos (2005) and Durante *et al.* (2013).

Having established robust evidence that individuals respect recipients' reference points over small dollar amounts, we then ask whether others' reference points help 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 accustomed to a certain level of income or consumption.

In the final part of the paper, we develop a survey experiment to test this claim. Subjects were 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 earning 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 an average income tax rate in the five-year scenario that is roughly 1.7 percentages points lower than in the one-year scenario (for which they chose a rate of 28 percent). This 1.7 percentage point difference is over half the size of the gap between tax rates chosen by Obama and Romney supporters, and is thus comparable in magnitude to other important predictors of tax preferences.[4]

Our work joins a large body of economics papers that attempts to incorporate insights from social psychology. Several studies have fruitfully modeled the implications of "behavioral agents" for the taxation of *goods*.[5] Some recent 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)).[6]

We add to this literature both in our approach, which involves lab and survey experiments rather than theory and calibration exercises, and our focus on reference dependence

[4]Interestingly, Gunnar Myrdal, in his voluminous study of U.S. race relations in the 1940s, found that white Southerners used reference-point arguments to justify why they need not redistribute to (much poorer) blacks. Whites would argue that they could never be expected to live on the same income as a black family because the cost of the lifestyle to which they were accustomed was higher: "People accustomed to suffer from want do not feel poverty so much as if they had seen better days....'People who have seen better days' are believed to be worse off than other paupers."

[5]See Bernheim and Rangel (2004) on optimal taxation when agents can become addicted to a good, O'Donoghue and Rabin (2006) on optimal "sin taxes" when agents lack self-control, and Allcott *et al.* (2014) on optimal energy taxes and subsidies when agents are inattentive. Recently, Lockwood and Taubinsky (2015) have re-examined sin taxes in settings with high wealth inequality, as such taxes tend to be regressive.

[6]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).

rather than other "behavioral" traits. We also ask a positive question—do respondents take reference points into account when acting as social planners—and for the most part leave aside the normative question of what they should do. While past work has explored whether reference dependence can help explain labor supply and buy/sell behavior of investors, to our knowledge we are the first to explore its ability to predict redistributive decisions both in laboratory settings and in preferences over real world policies such as income tax rates.[7]

We also contribute to a relatively recent literature on reconciling differences between individuals' stated policy preferences and the prescriptions from standard utilitarian optimal tax models. Both Saez and Stantcheva (2016) and Weinzierl (2014) point to individuals' reluctance to fully redistribute income *even in the absence of moral hazard* as a serious challenge to the utilitarian framework. We offer some evidence that the full-redistribution result can be largely rehabilitated within standard utilitarianism, albeit with non-standard individual utility functions. Both these papers note other important departures from utilitarianism (e.g., limited use of tagging) that we do not address in this paper.

Finally, to our knowledge, our study is one of the few to document choices that reflect the presumption of non-standard preferences *on behalf of others*. In this sense, our framework is related to models of guilt aversion (Battigalli and Dufwenberg, 2007), where individuals suffer disutility from a failure to meet the expectations of others, and as a result incorporate others' utility into their own choices (see Battigalli *et al.* (2013) and Ellingsen *et al.* (2010) for experimental tests of guilt aversion). Our results are novel in their application to important questions in public finance and their contribution to our understanding of redistributive preferences and policies.

The paper proceeds as follows. Section 2 presents a simple framework for thinking about optimal tax decisions with reference-dependent individuals. Section 3 describes the main lab experiment and Section 4 our mTurk sample. Section 5 reports the main results from the redistribution experiment. Section 6 introduces the tax-policy survey experiment and reports the results. Section 7 offers directions for future work and some concluding thoughts.

[7]For recent work on labor supply and reference points, see Fehr and Goette (2007), Farber (2008) and Exley and Terry (2016). For work on investor behavior and reference points, see, e.g., Genesove and Mayer (2001) and the finance literature on the disposition effect. Finally, Alesina and Passarelli (2014) analyzes how loss aversion affects policy formulation in general. Their finding of a status quo bias is driven by the strength of preferences by losers versus winners, rather than voters' regard for the losses of others, as in our approach.

2 Reference points and optimal taxation

We begin by briefly examining how reference-dependent utility affects the prescriptions of a standard model of optimal taxation, focusing primarily on a single-period static case. We close with a discussion of how dynamics and possible adaptation would change the analysis.

2.1 Redistribution in a static optimal tax model

2.1.1 With standard utility assumptions

Even in the absence of moral hazard, the standard optimal tax exercise would be intractable without assumptions on individual utility. As Mankiw *et al.* (2009) point out, despite the general caveat that utility cannot be compared across individuals, the typical optimal tax exercise does exactly that and further assumes that individuals have identical utility functions: a function $f : x \to u$ maps consumption x (which in the static model is identical to income or wealth) into utility u, with $f' > 0$ and $f'' < 0$.

With these assumptions, optimal taxation in the utilitarian framework with no moral hazard is a simple exercise. Assuming a single convex function can capture all agents' utility, welfare is maximized at the point of total equality.

Figure 1 illustrates this logic for two individuals p (poor) and r (rich), with ex-ante unequal wealth endowments x_p and x_r, $x_p < x_r$. Given r's lower marginal utility of wealth, a social planner can increase total utility by transferring some Δ from r to p. The social planner can continue to do so and still increase total welfare up to the point where $f'(x_r) = f'(x_p)$, that is, $x_r = x_p$.

2.1.2 With reference-dependent utility

If utility is reference-dependent, we lose the analytical convenience of both a common utility function across individuals and differentiability. Figure 2 shows typical reference-dependent utility functions.[8] Importantly, r and p no longer share a utility function, as each has a point of non-differentiability at their status-quo position. Transferring Δ from r to p no longer guarantees welfare improvements. As shown in the figure, moving to complete equality is welfare-reducing, as is any smaller perturbation that transfers endowments from r to p.

While we have drawn gains as concave and losses as convex in Figure 2, such restrictions are not necessary for the local result. So long as losses loom larger than gains relative to the reference point, then small amounts of redistribution will reduce welfare. It is harder

[8]We draw the shape of the utility functions in Figure 2 to roughly approximate those calibrated in Abdellaoui *et al.* (2007).

to make specific claims about much larger changes, which will depend on third derivatives. The main point is that reference dependence weakens the claim that, without moral hazard, redistribution is necessarily welfare-enhancing in the standard Mirrlees framework with a utilitarian social welfare function.

An alternative mechanism that delivers the same result is to assume that subjects have standard utility, but that the *social planner himself* overweights the utility of losers relative to winners, as in Saez and Stantcheva (2016). That is, loss aversion enters via the social welfare weights, not the individuals' utility functions. Our experiments will not be able to distinguish between these two mechanisms.

2.2 Reference points in a dynamic setting

The analysis in the previous subsection assumes a static setting where an individual forms a single reference point from her ex-ante endowment and then experiences a one-time change in utility based on the ex-post distribution. In reality, the government sets tax and transfer policy continuously. While a decision to substantially redistribute income or wealth in a given year might well have the utility consequences described above in the year that follows, if individuals adapt to their new reference points then the formerly rich will experience only momentary disutility in subsequent years.

It is an open question how quickly people adapt to changes in income and thus how malleable reference points are over time.[9] Strictly speaking, our empirical work will sidestep this question: we merely test whether individuals appear to respect others' reference points when deciding whether to redistribute. Finding, as we do, that they respect others' reference points implies that *individuals think* that others' reference points are not completely malleable, or that they think myopically when making redistributive decisions. Thus, the *actual* adaptation of reference points is not relevant to the positive question of whether individuals, acting as social planners, believe others' reference points should enter into redistributive decisions.

However, while not the focus of our analysis here, adaptation is an important consideration in drawing normative implications from this positive result. If adaptation is slow and losses loom larger than gains, then voters who seek to limit redistribution are plausibly maximizing welfare in the utilitarian model. While past work (Chetty and Szeidl, 2007) has shown that reference points (in their case, built on the micro-foundation of consumption commitments) imply higher levels of social insurance against *adverse events* (as losses have

[9]See, e.g., Di Tella *et al.* (2010) and citations therein. They find that individual happiness measures return to baseline roughly four years after an income shock.

greater utility cost with reference points than with standard utility), by the same logic reference points might also suggest lower levels of redistribution for the purpose of *condensing the current income or wage distribution*.

On the other hand, if adaptation is rapid, voters seeking to limit redistribution out of respect for others' references points might not be maximizing welfare, or at least not in a longer-run, steady-state sense. If voters or policy-makers overestimate the persistence of reference points, then their chosen level of redistribution may be lower than the optimal, welfare-maximizing point.

3 Experimental design

We recruited and compensated our subjects through Amazon's Mechanical Turk (mTurk) marketplace (which we describe in detail in Section 4), but redirect them to surveys that we built with Qualtrics' online survey software, adding functionality with JavaScript as needed.

We collected data from ten distinct sessions (and followed up with participants in one of these sessions one year later). In seven of these, respondents proceeded through modules of the survey in the following order: (1) the main redistribution experiment; (2) questions on loss aversion; (3) background questions on political beliefs and demographics. We describe each below. In three of the later surveys, subjects were presented first with questions on income taxation, which we describe in detail in Section 6. See Appendix Table 1 for the dates of the surveys and the attributes of each.[10]

3.1 Main redistribution experiment

The centerpiece of the survey presents respondents (whom we term "redistributors" in the paper, though at no time do we use this term in the survey itself) with the opportunity, in most cases hypothetical, to transfer money between two other anonymous mTurk 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.

[10]We had technical problems in one session and thus do not include it in the main analysis. A description of the problem as well as results from that session appear on the last page of the Appendix.

> 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.

To test the robustness of the results to within-person instead of between-person variation, we also performed the "reverse experiment" and so immediately after answering the treatment (control) version of the question, treatment (control) respondents answer the control (treatment) version of the question (with the labels "Persons A and B" replaced with "Persons C and D"). The survey is available for interested readers to take online.[11]

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 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,

[11] A version of the online experiment where the redistribution experiment is presented first can be taken at the following link: `https://az1.qualtrics.com/SE/?SID=SV_b2Tk5a7LuYAk38F&Preview=Survey&BrandID=columbia`. We provide the link to the tax survey experiment when we introduce it later in the paper. These versions of the surveys are in "preview" mode and thus we do not collect any identifying data—such as IP addresses—via these links.

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.

To ensure the robustness of our findings, we ran several variants on the main experiment described above. First, for one group of subjects (session 7), the subjects were informed, prior to seeing the instructions, that there was a ten percent chance their decision would be implemented for real stakes:

> The next two questions will give you the opportunity to determine the payments to two other mTurk participants.
>
> After you make your decisions, the computer will pick at random whether or not to implement your decision. There is a 10% percent chance that one of your decisions *will actually be implemented*. Because you do not know ahead of time whether your decision will be chosen, *you should make your decisions as if they were for real money*.

In another robustness check (session 8), none of the text was italicized or underlined, and the underlined reminder message (see screenshots in the Appendix) placed next to the slider was removed. In a final check (session 10), 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: <u>"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."</u> All other text was unchanged.

3.2 Questions to determine respondents' own sensitivity to reference points

As noted in the introduction, one of our initial goals was to examine whether redistributors' own loss aversion impacts the treatment effect. This analysis requires an individual-level measure of loss aversion. However, the vast majority of papers on loss aversion use between-subject analysis—for example, the classic endowment-effect experiments are demonstrated by one group's willingness to accept being higher than another group's willingness to pay which, by construction, is a between-subject exercise. Similarly, in their work showing that respondents judge the fairness of market transactions based on reference points, Kahneman *et al.* (1986) use comparisons between groups assigned to read different vignettes.

Our approach is largely to take the questions asked in between-subject designs and present both versions to each subject. We provide complete descriptions of these questions in the Appendix. To briefly summarize, in one set of questions, respondents face hypothetical lotteries of increasing risk; we use the number of lotteries they choose to play as a measure of risk-seeking. Subjects do this exercise for lotteries in both the gain- and loss- domains, allowing us a rough comparison of the shape of the utility function above and below the status quo reference point. In a second measure inspired by Genesove and Mayer (2001), we ask respondents for the lowest price they would accept in selling a hypothetical house under two scenarios: (a) if the house was purchased for $300,000, and (b) if it was purchased for $250,000. We treat the difference as a measure of reference dependence. Third, inspired by Kahneman *et al.* (1986), we ask respondents to evaluate the fairness (on a scale of one to four) of cutting the wage rate of an existing employee versus offering a lower wage to a new employee, again treating the difference as a measure of respect for reference points.

Finally, because of concerns that asking own loss aversion questions immediately after completion of the main experiment might create biases that are hard to sign *a priori*, we followed up with subjects from session 3 of the survey two years later. In this survey, no experiment was performed, and we only asked loss aversion and demographic questions. Besides the above loss-aversion questions, we also added a loss aversion question inspired by Carpenter *et al.* (2005), asking for willingness to sell versus willingness to buy the same object, treating the difference as a measure of reference dependence. Again, we present the exact wording in the Appendix.

3.3 Demographic and political opinion questions

The survey ended with standard demographic questions, as well as a question of affiliation in the 2012 presidential election. These questions allow us to examine whether our treatment effect is larger for certain groups, and also to compare the mTurk sample to more representative populations such as the American Community Survey. Given that we collect these questions *after* the experiment (so as not to prime the results in the experiment), it is possible that some answers may be primed by the experiment itself; comparisons by these covariates should thus be viewed with this potential priming in mind. 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.

4 Data

4.1 Data collection procedures

All of our subjects were recruited through mTurk, an online labor market where "requesters" can post *human intelligence tasks* (HITs) to be completed by "workers." As of the summer of 2014 (when much of our data were collected), mTurk advertised that requesters could "access more than 500,000 workers." The most common posted HIT at that time was "extract purchased items from a shopping receipt" and paid 8 cents (the requester also pays a ten percent commission to Amazon).[12]

Over the past few years, social scientists have increasingly used mTurk to perform experiments and collect survey data (see Kuziemko *et al.*, 2015 and papers cited therein for a review). 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, which approximated minimum wage assuming that subjects took seven or eight 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, reading through worker forums suggests that we are paying a very generous wage (and indeed when we post a request for 300 survey takers, the full sample is typically gathered within an hour).

Each worker logs in with an mTurk worker ID. We collected data over ten separate sessions, dropping any worker who has taken a previous survey with the same ID so as to gather a fresh sample each time (though our main results hold when we keep repeat-takers in the sample, see Appendix Table 7). Of course, if workers have multiple worker IDs then some individuals may have participated in a previous session. Outside of surveys (which appear to make up a very small share of all HITs), in which case requesters may want unique workers, 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, and thus could have passed through our screening process.

Another issue that arises on mTurk is the possibility of "robots," algorithms that masquerade as humans. To address this concern we begin each survey with a "captcha" (non-standard writing difficult for computers to interpret).[13] In fact, when we ask respondents

[12]Details cited in this paragraph are based on viewing the mTurk homepage on 10:56 AM EDT, August 12, 2014. As of July 5th 2016, mTurk advertised access to more than one million workers.

[13]Examples of "captchas" can be found here: `http://www.fileflash.com/graphics/screens/Captcha_Creator_PHP_Script-69.gif`.

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 may be learning how to read these "captchas," one of the authors drew her own sketches of a cat, dog, horse and panda bear, and respondents in some later rounds were asked to answer multiple-choice questions of the form: "this is a picture of...." after seeing these sketches. To the extent that some robots still remain in our sample after these checks, they would attenuate any treatment effect.

To limit heterogeneity of the sample and to keep background conditions as similar as possible across surveys, we collect all data on workdays during daylight hours on the East Coast of the United States. Individuals were automatically prompted for a response when they tried to skip questions (to discourage robots or inattentive respondents). Particularly given our focus in some parts of the survey on American tax policy, we limited the survey's availability to those with U.S. billing addresses; we further asked respondents to confirm their residency in the United States. To further ensure the attentiveness of our subjects, we limit respondents to those with positive ratings from at least 90 percent of past requesters.

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.[14]

In Appendix Tables 2 we show how questions on perceived political bias of the survey vary with treatment status. About 88 percent of respondents felt that the survey was unbiased, with about eight percent finding that it had a liberal slant and three percent a conservative slant. None of these three measures is different across treatment and control. 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.

4.2 Data sample

Table 1 provides detail on the mTurk workers who completed our survey, comparing them to the (weighted) population of adults sampled in the 2010 American Community Survey. Consistent with past work using mTurk, we find that younger, male, and college-educated subjects are over-represented in our sample, while minorities are under-represented.

Table 2 provides a longer list of covariates, while Table 3 examines differences between the

[14] We suspect that the positive feedback likely reflects the tedium of most other mTurk tasks. As noted, essentially all of the negative feedback concerned the difficulty of some of the "captchas."

control and treatment groups. Overall, there appears to be good experimental balance, with no variable showing a statistically significant difference at the five-percent level. In particular, a variable that would be expected to have an impact on redistributive decisions—an indicator for supporting President Obama in the 2012 election—is essentially identical between the treatment and controls groups (p-value of 0.89).

5 Results from the redistribution experiment

5.1 Main results

Table 4 shows, for the full sample, 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 complete redistribution. 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 \$4.55 from the richer recipient to the poorer one, or 91% of the level of redistribution required for strict equality. Recall that the default position of the slider was the status quo (\$5 and \$15) allocations, suggesting that anchoring may, if anything, bias the control group results against inequality-reducing redistribution. Those assigned to the treatment redistribute on average \$0.80 (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.[15] The control group mean increases slightly, consistent with the view that some rapid finishers simply clicked thoughtlessly through the redistributive decisions, leaving the sliders in their default positions. The treatment effect increases when these subjects are dropped.

Column (3) further excludes subjects for whom the \$5/\$15 experiment was not the first item of the survey. This restriction removes subjects that may be contaminated by exposure to our income tax survey experiment (which we discuss in Section 6 below). We take the results in this column as our preferred specification: the control group closes nearly 94% of the gap ($4.682 \div 5$), whereas the treatment group only 77 ($(4.683 - 0.829) \div 5$). Thus, our treatment reduces redistribution by 17 percentage points.

The rest of the table provides a number of additional robustness checks. In column (4) we drop subjects who choose to make inequality-*increasing* reallocations. We present results with these subjects excluded to ensure that our average treatment effect is not being driven

[15] Specifically, less than three minutes for the first session (as it did not have the module on loss aversion), and six minutes for the other sessions.

by them (though one could imagine such choices as welfare-maximizing under, say, a convex utility function). Column (5) includes a number of demographic control variables; given the balance across the control and treatment arms documented in Table 3, it is not surprising that the treatment effect is unchanged with the inclusion of these controls. Column (6) presents results for the "real stakes" subsample of redistributors who were informed that there was a 10 percent chance that their decision would be implemented, while in Column (7), we show results for the group of subjects where none of the text in the experiment was highlighted or underlined. We find that our basic result continues to hold in each of these subsamples, though the effect is somewhat attenuated, particularly in the non-underlined version.

Finally, in column (8) we present the results of the variant designed to assess whether procedural justice concerns can account for our treatment effect, labeled "Ex. promised payment" in the table. Recall that in this version, subjects in the treatment group were informed that "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." While the treatment effect is smaller than that observed in our main results (-0.51 versus -0.83), it is still significant at the 10% level. This check is quite demanding: by warning (hypothetical) subjects in the treatment condition that their ex-ante allocation could change, we are not only avoiding making a promise but also softening their reference points. As such, even if procedural justice issues played no role in the main treatment effect, we might expect the treatment effect in this session to be somewhat smaller if redistributors are instead responding to the softened reference point of the hypothetical subjects.

Our main results are consistent across sessions more generally, as shown in Figure 3. Here, using our preferred sample restrictions, we plot the coefficients and 95 percent confidence intervals for the treatment effect, disaggregated by session date. The only treatment effect that stands out is the very first survey, which is somewhat larger in magnitude than the others (though we cannot reject equality with all other rounds at even the ten percent level).[16] The session-by-session results highlight how the "real stakes" and "unemphasized" versions of the survey have nearly identical treatment effects as the two standard surveys that followed the initial session.

As noted in Section 3, all subjects are presented with the "reverse" experiment: those first assigned to the treatment also face the control scenario, and vice versa. Our emphasis

[16]While, as noted, we drop individuals who participated in previous sessions with the same mTurk ID, the larger treatment effect for the first survey potentially suggests that individuals who already took the survey using a different worker ID may have attenuated the measured treatment effect in later sessions. However, this interpretation is quite speculative given that the difference in effect across rounds is not statistically significant.

is on the between-subjects analysis presented above, given that respondents may anchor at least partially on their first response, and past work has further shown that the tendency to anchor may be related to loss aversion.[17] Nonetheless, the within-subject treatment effect is highly significant, and near-identical in magnitude to the between-subject estimates. Table 6 shows an average treatment effect of -0.873 (column 1), with a somewhat smaller effect (column 2) for those who start with the treatment scenario than those who start with the control scenario (column 3). The final column shows that the within-subject result holds also for the real-stakes session.

5.2 Intensive versus extensive margin effects

Figure 4 shows histograms of the final allocation for the ex-ante "poorer" player, both for the treatment and control groups. For both groups, the distribution is bimodal, with most of the mass at (10, 10) but also a second, shorter peak at (5,15); there is almost no mass in between these two points. Thus, most of the treatment effect occurs at the *extensive* margin—the decision to redistribute at all—as opposed to increases in partial redistribution. If losses are "more convex" than gains are "concave," then after no redistribution, complete redistribution is the utility-maximizing outcome. As such, the lack of partial redistribution is consistent with subjects responding to the convexity of losses.

Table 5 shows these extensive margin results in a regression framework. Approximately four-fifths of control group respondents set final allocations at (10, 10), as compared to only 60 percent of the treatment group, for a treatment effect of 25 percent along the extensive margin. As before, this result is highly robust to the addition of controls, as well as excluding suspiciously short surveys and choices that increase inequality.

5.3 Magnitude of the reference point effect

To gauge the magnitude of the effect we document above, we compare our treatment effect 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 using polling data, which has shown that one of the most important determinants of support for redistribution is whether income is seen as the result of merit versus luck (Alesina and Angeletos, 2005).

To make this comparison, we ran a session where the control arm was kept the same (the \$5 and \$15 endowments were determined by a coin flip, and the recipients will 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.

[17] See Beggs and Graddy (2009).

in original].” As with the control version of the experiment, the redistributor is told that Persons A and B would only learn of their own final allocation.

The results are reported in Table 7. Consistent with past work demonstrating the important role of perceived merit in attitudes toward redistribution, the coefficient on the treatment variable is large in magnitude, indeed larger than in our reference-point experiment. Dividing the coefficients in this table by their analogues in Table 4 suggests that our reference-point effect is about half of the luck-versus-merit difference.[18] It is possible that the SAT treatment contains an implicit reference point treatment as well: a recipient would likely have a rough sense of their performance on the test, and thus form a reference point over their expected compensation. To the extent that redistributors take this expectation into account, we would over-estimate the pure “merit” effect (thus underestimating the magnitude of our main reference-point treatment effect as a share of the merit effect).

Interestingly, the histogram of final outcomes in the luck-merit experiment takes a somewhat different shape than what we observe for the main experiment (see Appendix Figure 5). There are still mass points at $5 and $10, but other intermediate choices are now more common. Whereas, conditional on redistributing something, only 12 percent of those assigned to the original treatment chose an “in between” final allocation in the reference-point experiment, 28 percent do so in the luck-versus-merit experiment, highlighting that respondents are more sensitive to the assigned reference points in our main loss aversion treatment.

5.4 Treatment effect heterogeneity

If the recognition of loss-averse preferences among recipients explains the behavior of redistributors, it seems plausible that this effect will be stronger for subjects who are themselves loss averse, and hence are more likely to project such preferences onto others. (Recall that the description of our loss aversion measures are provided in the Appendix.) To explore this possibility, we now turn to examine how the treatment effect varies with measures of own loss aversion using regressions of the form

$$Redistribution_i = \beta_1 Loss\text{-}Averse_i \times Treated_i + \beta_2 Treated_i + \beta_3 Loss\text{-}Averse_i + \gamma X_i + \epsilon_i.$$

If subjects project their own loss aversion onto others, then $\beta_1 < 0$; that is, those who are loss-averse will redistribute relatively less than others when exposed to the treatment

[18]Our luck-versus-merit results are similar to those found in preliminary work by Chevanne et al. In an undated online draft of their paper, they find that third-party dictators redistribute roughly 32 percent less when they are told that initial inequality was due to effort, quite close to the effect we obtain in our experiment (a 36 percent reduction in the specification with controls).

condition.[19] We present the results in Table 8.

The results are mixed. Individuals who are risk-seeking in the loss domain are much more sensitive to the treatment: in fact, the main effect of treatment is rendered insignificant once this interaction term is included in the regression (col. 1). Interestingly, being risk-seeking has no effect on the size of the treatment effect (col. 2). The difference in risk-seeking in the loss domain relative to the gain domain is also highly significant (col. 3). While the point estimates on these interaction terms line up as we predicted, the coefficients on the interaction terms in cols. (5) and (6), while not statistically significant, suggest if anything that those who exhibit reference-dependence (as we measure it) were less responsive to the treatment.

As we observed earlier, one concern with the own loss aversion measures collected immediately after the $5/$15 experiment is that it may prime respondents in a manner that is hard to sign *ex ante*. For this reason, we recontacted our session three respondents 24 months after they took the original survey, and ask *only* the loss aversion and demographic questions (thus the follow-up survey contains no experiment). The follow-up response rate was 25%, which pleasantly surprised us given the long time lag. We show in Appendix Table 3 that, as expected, answers to questions on fixed demographic traits (e.g., gender) are identical in the original and the follow-up survey. Stated support for Obama in the 2012 election is nearly identical in the original and follow-up, and opinions about redistribution in general are also highly correlated across time. Interestingly, however, responses to the loss-aversion questions in the original survey do not predict how respondents answer the same question two years later in the follow-up survey.

As reported in Appendix Table 4, the main treatment effect (again, based on respondents' original answers in session three) remains significant in this sub-sample, with a point estimate essentially identical to the main results reported in Table 4, suggesting that along this important measure those who chose to take the follow-up survey are similar to the full sample. Again, however, we find no consistent evidence that own loss aversion helps predict respondents' treatment effect.[20]

The difficulty in interpreting differential treatment effects by measures of own-loss-aversion,

[19] This regression presents a demanding test: as Table 2 shows, a large majority of individuals give answers consistent with loss aversion in the surveys. Given that the main treatment effect acts primarily through the extensive margin, it is plausible that individuals with varying degrees of loss aversion over own payoffs will have similar treatment effects, so long as they are above some threshold level of loss aversion.

[20] While our recontact sample did not prove particularly illuminating, the ability to recontact mTurk respondents may be useful in many contexts. We are happy to share the scripts we used to implement the recontact survey to researchers interested in conducting follow-up surveys on mTurk.

however, is made apparent when we consider the surprisingly low correlation among our measures of subjects' own loss aversion, which we show in Appendix Table 5 for those loss aversion measures gathered contemporaneously with the \$5/\$15 experiment.[21] The correlations between any pair of loss aversion measures bounce between positive and negative, suggesting that they are capturing distinct phenomena. If this is the case, it is not surprising that they would then be uncorrelated with how one thinks about applying loss aversion in yet another distinct reference dependence domain, involving consideration of the endowments of others.[22]

Beyond measures of reference-dependence, we also examine whether the magnitude of the treatment effect depends on other demographic and background characteristics. Quite surprisingly given our large sample size, of the eleven variables we interact with the treatment dummy, only one interaction (age) is even marginally significant (see Appendix Table 6). As such, differential treatment effects appear remarkably small in general for our treatment.

Overall, the results in this section do not allow us to link subject attributes to the strength of the treatment effect, which might in turn have allowed us to further focus the channel through which the reference-point frame affects redistribution. At the same time, the lack of heterogeneity across subjects emphasizes the robustness of the reference-point treatment to individual attributes.

5.5 Discussion

We find very robust evidence that in their role as social planner, subjects' redistributive decisions are affected by whether recipients are aware of their initial endowments. There are several primary candidate explanations for this treatment effect. Redistributors may (1) respect property rights over initial endowments, (2) exhibit standard status quo bias, (3) view initial endowments as a commitment or promise to recipients that they do not wish to break, and (4) assume recipients have reference dependent preferences. The data are harder to reconcile with either of the first two explanations. The most straightforward property rights explanation is inconsistent with the near-complete redistribution in our control condition. This result also casts doubt on pure status quo bias, which would also limit redistribution

[21]Correlations (available upon request) across questions are not any more positive in the small recontact sample.

[22]We had initially expected that individual measures of loss aversion would play a larger role in our study. Our attempts to adapt what past authors have used and to develop our own measures were not overwhelmingly successful, as we would have expected them to better correlate with each other. Developing individual-level measures of loss aversion or reference-dependence that are both easy for respondents to understand and predictive across different domains seems like a useful area on which future work could focus.

in the control scenario.[23]

Of course, applying the notion of "property rights" is complicated in our control condition: can a recipients have "property rights" over claims that they are unaware that they possessed? In the "no promises" variant of the experiment, however, "property rights" in the treatment condition are deliberately weakened even further by warning the recipients that their ex-ante allocations could change. Yet the treatment effect remains, suggesting a respect for even a broad sense of property rights is not the main constraint on redistributors in the treatment condition. Moreover, this "no promises" variant also argues against a dominant role of distributive obligations as a result of commitments or promises.

Thus, while we do not have decisive evidence to adjudicate among explanations, we argue that recognition of others' loss aversion is most easily reconciled with our full set of results.

These experimental results suggest that respect for reference points may help to explain why individuals eschew complete redistribution even in the absence of moral hazard, in contradiction to the prescriptions of the optimal tax model. When redistributors need not consider recipients' reference points, as in our control condition, we essentially recover the Mirrleesian full-redistribution result, making the results from the control arm interesting in their own right. We can thus rehabilitate the full-redistribution result within classic utilitarianism, albeit with non-standard utility functions.

As noted earlier, Chetty and Szeidl (2007) present a model of consumption commitments that could similarly diminish redistribution by a social planner who takes into account the commitments of relatively well-off individuals. As they observe, however, in a context such as ours it is implausible that *actual* consumption commitments could drive subjects' decisions—the individuals over whom they were making decisions were given money that, by construction, had not yet been spent.[24]

Our findings to this point suggest that subjects are sensitive to others' reference points in redistributive decisions in laboratory settings over small stakes. In order 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.

[23]See Trump (2015) for experimental work on how respondents might deem high levels of inequality legitimate because of a belief in a "just world."

[24]If individuals develop reference-dependent heuristics as a result of commonly observing consumption commitments in their day-to-day lives, it could help to provide an underlying model for reference dependent preferences. Examining this possibility may be an interesting direction to pursue but is outside the scope of our paper.

6 Survey results on reference points and preferred top income tax rates

The question of what constitutes an appropriate income tax rate on high-income households is a much-discussed issue in American politics today. A threshold of $250,000 has become a focal point in this discussion, and surveys often ask about support for higher taxes on households with annual incomes of at least that level.[25] While the majority of respondents in surveys tend to support higher income taxes on this group, the strength of those preferences has been debated. For example, while survey respondents in 2010 exhibited strong support for letting the so-called "Bush tax cuts" (those specified in the 2001 and 2003 tax relief acts) expire for individuals earning over $250,000, in that year's midterm Congressional elections Republicans won handily despite their having made extending these tax cuts a major focus of their campaign.[26]

The survey experiment below tests whether respect for reference points might weaken respondents' preference to tax high-income households.

6.1 The survey experiment

Subjects were presented with a vignette describing an individual that had received an unexpected increase in earnings. In most waves, the source of the increase was a corporate takeover of the company where the individual is employed (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. [27]

The "control" arm of the vignette took the following form:[28]

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

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

[26] Larry Bartels discussed this tension in a 2010 online post: `http://today.yougov.com/news/2010/10/26/taxes-energized-minority/`.

[27] Interested readers can take the survey themselves. A version where the tax experiment is presented first can be taken at this link: `https://az1.qualtrics.com/SE/?SID=SV_0MnchCiPWRxAsqV&Preview=Survey&BrandID=columbia`. This version is in "preview" mode and thus we do not collect any identifying data—such as IP addresses—via this link.

[28] In the vignettes we reference a tax rate of 22 percent on the "average American." We base this figure on NBER Taxsim estimates for combined federal and state income tax, and then add the employee side of payroll 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.)

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, to assess the generalizability of our findings, we changed the reason for the individual's increase in income. In the control version of this later wave, 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, to strengthen the reference point, we simply replaced "*This year*" with "*Five years ago*" and changed the verb tense (from "will receive" to "receives") as appropriate.

In a subset of sessions, following the vignette and the associated slider question, we asked subjects about their tax preferences more generally as follows: "In general, how do you feel about increasing taxes on those making $250,000 or more (as has been proposed in Congress recently)?" Respondents could choose from "strongly oppose," "oppose," "favor," and "strongly favor" (though we collapse these choices into an indicator variable for favoring or strongly favoring the policy).

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 nearly five years will set a lower ideal tax rate than those presented with a protagonist that has received high earnings for only a short time, and hence is not yet accustomed to it.

6.2 Results

We begin by presenting results based on the pooled sample of both takeover and lottery vignettes. In the first column of Table 9, 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 for the protagonist that is 1.17 percentage points lower than control subjects, significant at the 10 percent level; by comparison, the control group mean is 28.8 percent. As with the $5/$15 experiment, our preferred sample includes only those who saw the tax experiment first (and thus cannot be contaminated with the $5/$15 redistribution experiment). When we focus on the subsample where the tax vignette appeared first (column 2), the treatment effect increases to 1.71 percentage points (with a control group mean of 28.6).

A small fraction of subjects choose extreme values: about one percent of subjects selected a tax rate of zero while a few chose tax rates of 99 and 100 percent. In column (3) we omit the top 1 percent and bottom 5 percent of observations to limit the influence of these extreme observations.[29] This restriction has only a slight impact on the size of the treatment effect. In Appendix Table 8, we show that the treatment effect is robust to a number of alternative ways of dealing with outliers, including estimates based on median regressions, winsorizing instead of dropping outliers, dropping only zero tax rates, and dropping regressive (<22 percent) tax rates. In all specifications, the treatment effect's magnitudes are comparable to the figures presented in Table 9. In column (4) we include controls, which has essentially no

[29]Another reason to exclude zero in particular is that it is the default position of the slider and thus many of these individuals may have been simply skipping through the survey.

impact on the treatment effect.

While we tried to hold everything constant in the treatment and control arms except the strength of the reference point, it is possible respondents read other differences into the stories. We suspect that the 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 scenario the protagonists enjoys the large raise after having barely worked for the company, whereas in the one-year scenario he put in his time before getting the big raise. Given the greater willingness to redistribute gains due to luck both in our \$5/\$15 experiment as well as in work cited earlier, respondents should choose a higher tax rate for the protagonist in the control scenario.

However, a concern that pushes in the opposite direction in the takeover vignette is that individuals confronted with the five-year scenario may credit the protagonist with greater merit because he has worked at the larger 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* 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 half a decade. In the one-year scenario, the future performance of the protagonist at the new corporation is left unclear.

For this reason, it is useful to examine the estimates from the takeover and lottery vignettes separately, as only in the former case would the merit argument apply. Column (5) shows estimates from the takeover vignette and column (6) the lottery vignette. While the treatment effect for the takeover vignette is larger (-0.0185 versus -0.012 for the lottery vignette, the latter of which is not statistically significant), both are negative and are statistically indistinguishable from one another. The preferred *levels* of taxation in the lottery vignette are not higher than in the takeover vignette, a somewhat surprising result given that respondents typically redistribute less when income is due to merit rather than luck, and presumably working at a company that does well would be considered a more "deserving" situation than winning the lottery.

In the final two columns, we test whether the treatment moves respondents' views about upper-income tax rates more generally. Treatment respondents are seven percentage points (roughly ten percent) less likely to support a tax hike on individuals making \$250,000 or more, though significant only at the ten-percent level. This result suggests that the vignette's framing is sufficiently salient to affect respondents' views on preferred tax rates for *all* high-income individuals, not just the one portrayed in the vignette. Of course, this outcome may be primed by the answer to the slider question (which always precedes it), so should be interpreted with caution.

6.3 Discussion

The results from the survey experiment show that individuals appear to reward more deeply embedded references points (i.e., income levels that have been experienced for longer periods of time) with lower tax rates. The magnitude of this effect is quite large. For example, Obama supporters 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."[30]

While the survey experiments have documented that the strength of an individual's reference point reduces the tax rates assigned by our subjects, the precise mechanism is unclear. 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 could well be responding to the perceived consumption commitments of the protagonists in the vignettes, in the spirit of Chetty and Szeidl (2007). It is plausible 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 amongst underlying explanations for this behavior.

It is also interesting to note that, at least along some policy dimensions, there appears to be respect for reference points in the distribution of *transfer* policies as well. For example, a policy that has gained popularity during the Great Recession would require welfare and food stamp recipients to pass drug tests. In a 2011 Rasumussen poll, while 53% of respondents supported mandatory drug test for *new applicants* to welfare, only 29% supported that same requirement for *current recipients*.[31] As such, respondents seemed to view individuals' current benefits as more of an entitlement. The fact that many cuts in benefits are "grandfathered in" can also be viewed as an implicit respect for beneficiaries' reference points.

[30]To estimate this effect, we use our preferred specification from Table 9 (i.e., column 3) but substitute a "Supported Obama" indicator for the treatment indicator. The mean tax rate chosen among the control group (i.e., those who did not support Obama) is 26.1 percent. When we instead repeat this analysis with the binary "support a general upper-income tax hike" variable as our outcome, our treatment effect is 29% of the corresponding "Obama" effect.

[31]See `http://stopthedrugwar.org/chronicle/2011/jul/22/national_poll_finds_support_welf` for a discussion of these results.

7 Conclusion

Past work has established that in many contexts, individuals display reference-dependent preferences. We provide robust evidence that, in a laboratory setting, 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 close to the full ex-ante income gap. However, redistribution is reduced by nearly twenty percent when the recipients *do* know their ex ante endowments. This reference-point effect is large in magnitude, more than one-half of the effect of having endowments determined via performance on an academic test versus a coin flip.

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.

Future work might examine the extent to which preference anomalies exist in decisions with direct consequences for others' (rather than one's own) payoffs in contexts other than redistribution. The limited work we have found on this question suggests that context may be important, as Marshall *et al.* (1986) do not find, as we do, that individuals project loss aversion when they act as *advisors* on others' behalf and Andersson *et al.* (2014) find that loss-aversion is muted when deciding over risky gambles that affect others rather than oneself. Beyond voters and advisors, there are many other roles that entail making decisions with consequences for others, such as parents, employers, and physicians.

In this paper, we have not confronted the normative issue of whether preference anomalies such as loss aversion *should* be respected by the social planner. Suppose that individuals behave as if their own utility is reference-dependent and that, as our work shows, they take others' loss aversion into account when acting as social planners. Should policy-makers and economists nonetheless still restrict themselves to classical utility and social welfare functions when conducting welfare analysis? Such questions have recently been taken up in Lockwood and Weinzierl (2016).

The existence of reference-dependent preferences in redistributive decisions may also help to explain some puzzling aspects of tax policy. For example, if wealth (a stock) is a more salient reference point than income (a flow), it could help to explain the lack of broad-based support for wealth taxes (and may be reason for skepticism that recent wealth tax proposals will get much traction). Our findings may suggest that tax increases—whether based on wealth or income—might be better-received if policymakers can commit to them several years in advance of their implementation, thus allowing individuals to adjust their reference wealth or income ahead of the actual change. A fuller analysis of the consequences of reference-dependent utility for taxation—how reference points are set and evolve in response to policy changes or pronouncements; whether there are circumstances that attenuate or intensify the role of reference-dependence in redistributive preferences; and so forth—is a further area for future research.

References

Abdellaoui, M., Bleichrodt, H. and Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, **53** (10), 1659–1674.

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

— and Passarelli, F. (2014). Loss aversion and politics.

Allcott, H., Mullainathan, S. and Taubinsky, D. (2014). Energy policy with externalities and internalities. *Journal of Public Economics*.

Andersson, O., Holm, H. J., Tyran, J.-R. and Wengström, E. (2014). Deciding for others reduces loss aversion. *Management Science*, **62** (1), 29–36.

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

Battigalli, P., Charness, G. and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, **93**, 227–232.

— and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review*, pp. 170–176.

Beggs, A. and Graddy, K. (2009). Anchoring effects: Evidence from art auctions. *The American Economic Review*, pp. 1027–1039.

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.

Carpenter, J., Verhoogen, E. and Burks, S. (2005). The effect of stakes in distribution experiments. *Economics Letters*, **86** (3), 393–398.

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

Di Tella, R., Haisken-De New, J. and MacCulloch, R. (2010). Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior & Organization*, **76** (3), 834–852.

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.*

Ellingsen, T., Johannesson, M., Tjøtta, S. and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, **68** (1), 95–107.

Exley, C. L. and Terry, S. J. (2016). Wage elasticities in working and volunteering: The role of reference points in a laboratory study.

Farber, H. S. (2008). Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *The American Economic Review*, **98** (3), 1069–1082.

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

Fehr, E. and Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *The American Economic Review*, **97** (1), 298–317.

Genesove, D. and Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, **116** (4), 1233–1260.

Harms, P. and Zink, S. (2003). Limits to redistribution in a democracy: a survey. *European Journal of Political Economy*, **19** (4), 651–668.

Kahneman, D., Knetsch, J. L. and Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review*, pp. 728–741.

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.

— and Taubinsky, D. (2015). Regressive sin taxes.

Lockwood, B. B. and Weinzierl, M. (2016). Positive and normative judgments implicit in us tax policy, and the costs of unequal growth and recessions. *Journal of Monetary Economics*, **77**, 30–47.

Mankiw, N. G., Weinzierl, M. and Yagan, D. (2009). Optimal taxation in theory and practice. *The Journal of Economic Perspectives*, pp. 147–174.

Marshall, J. D., Knetsch, J. L. and Sinden, J. A. (1986). Agents' evaluations and the disparity in measures of economic loss. *Journal of Economic Behavior & Organization*, **7** (2), 115–127.

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

MIRRLEES, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, **38** (2), 175–208.

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.

ROEMER, J. E. (1998). Why the poor do not expropriate the rich: An old argument in new garb. *Journal of Public Economics*, **70** (3), 399–424.

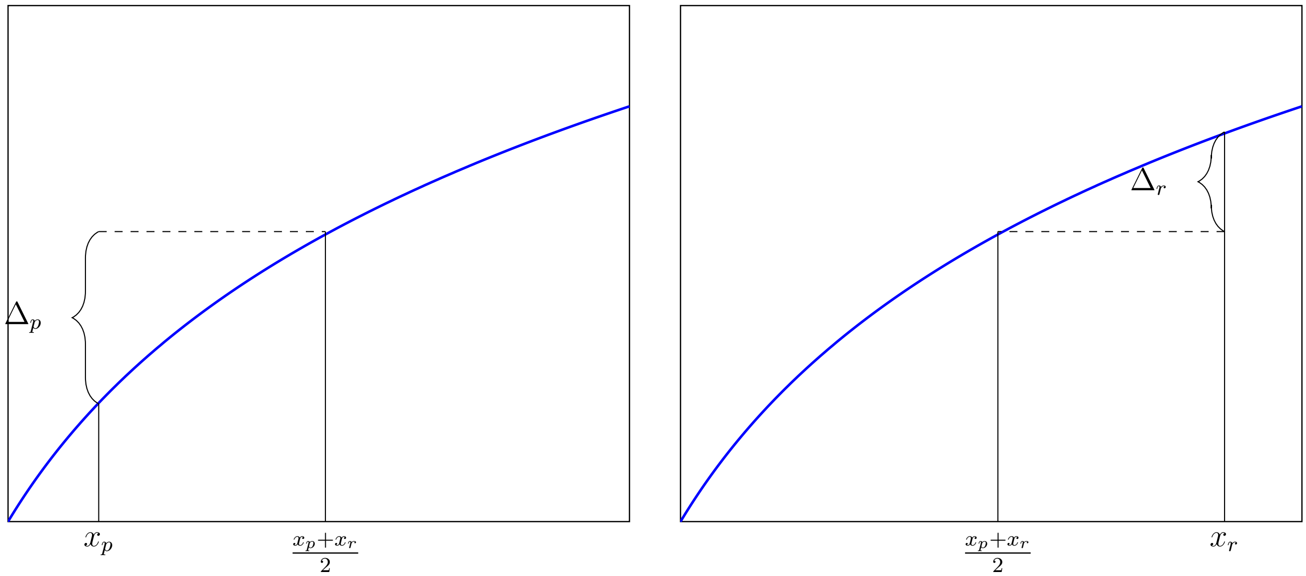
SAEZ, E. and STANTCHEVA, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *The American Economic Review*, **106** (1), 24–45.

TRUMP, K.-S. (2015). Accepting inequality: How what is influences what ought to be.

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 standard utility functions

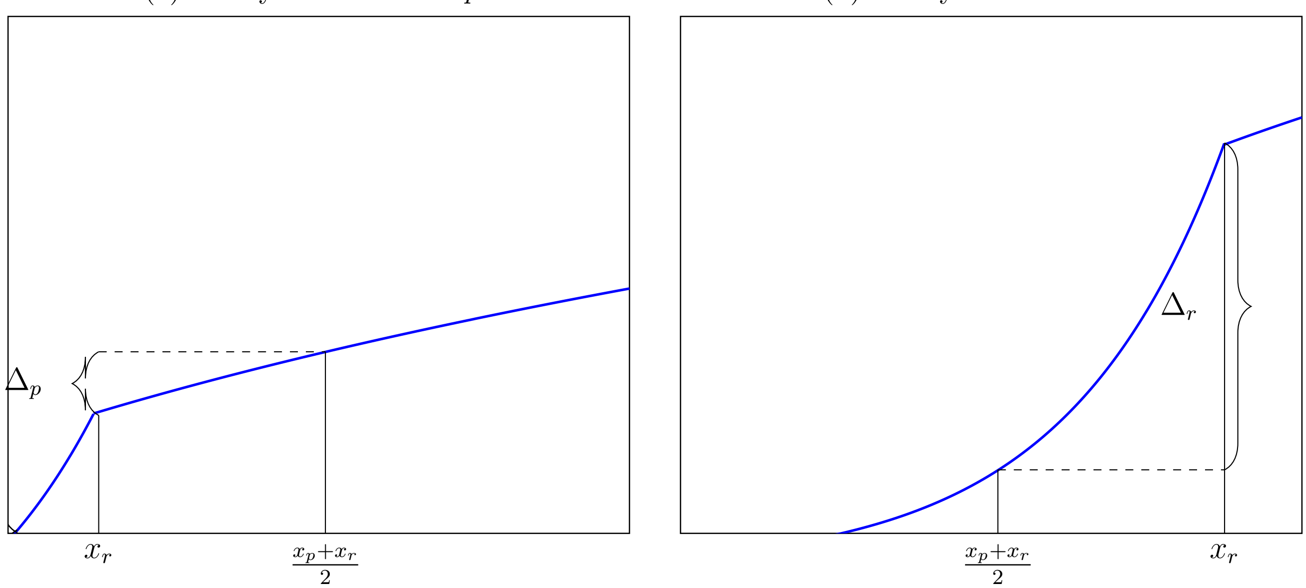
(a) Utility for individual p (b) Utility for individual r



Notes: A depiction of the optimal tax solution under a utilitarian social welfare function when utility (y-axis) is a positive and strictly concave function of consumption (x-axis).

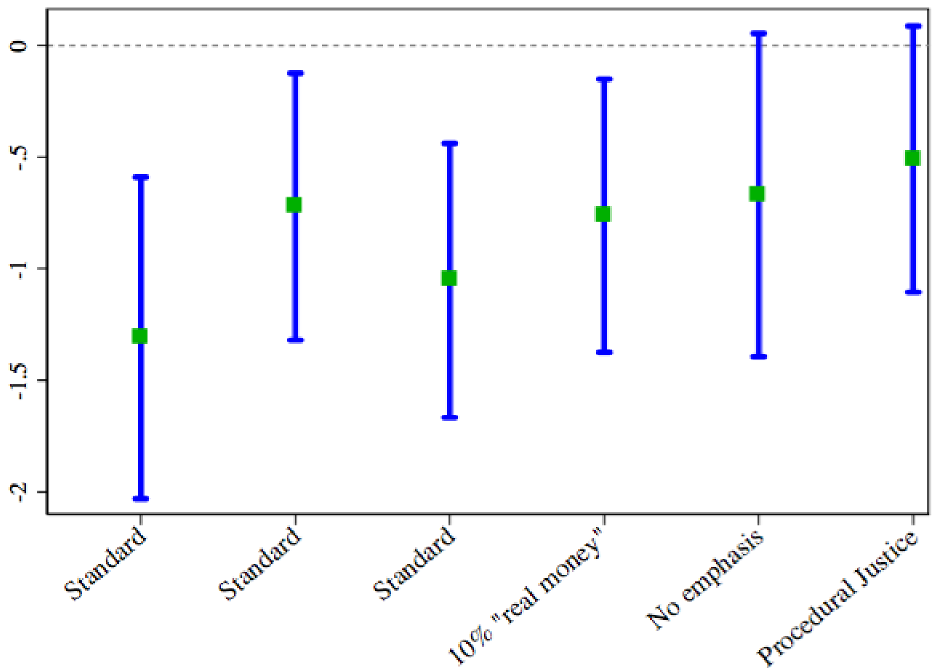
Figure 2: Redistribution with reference-dependent utility

(a) Utility for individual p (b) Utility for individual r



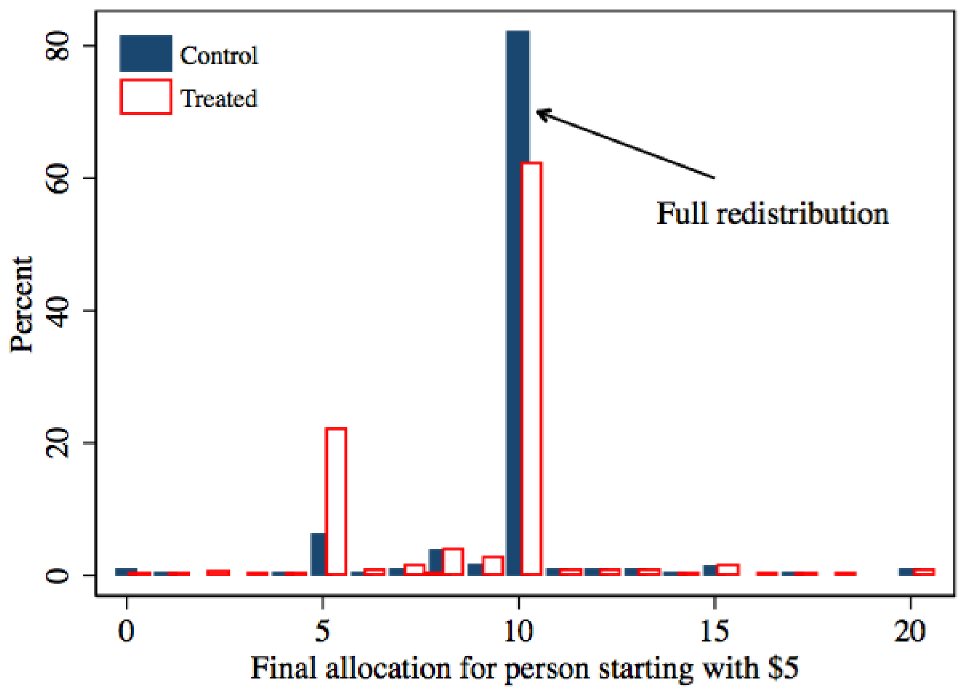
Notes: A depiction of changes in utility after full redistribution when utility functions exhibit loss aversion. In this example, we have drawn losses from the reference point as convex and gains as concave.

Figure 3: Treatment effects and ninety-five-percent confidence intervals for the reference-point experiment, by session



Notes: We show separately for each session the between-subject treatment effects for all rounds in which the reference-point money-transfer experiment appears first (the sample in col. 3 of Table 4). As noted in the text, there are a total of seven sessions where the money-transfer experiment appears first, but one (session six) contrasts endowments gained by luck versus merit instead of references points. Those results are reported in Table 7 but are not plotted here.

Figure 4: Histogram of ex-post allocations for the ex-ante poorer player



Notes: Sample used in the figure is that in our “preferred” analysis sample (Column 3, Table 4).

Table 1: Basic summary statistics in mTurk sample compared to ACS

	(1) mTurk Sample	(2) ACS sample
Age	33.04	46.40
Female	0.444	0.515
White	0.773	0.669
Black	0.0730	0.120
Hispanic	0.0563	0.143
Asian	0.0764	0.0503
College	0.449	0.257
Income ($1,000)	49.47	71.32
Obervations	2,041	2,369,395

Notes: Observation totals are the shared non-missing observations across all variables. Col. 1 includes all ten sessions of the experiment. Col. 2 includes all adults in the 2010 American Community Survey (weighted with the provided individual-level weights). "Income" refers to household income (in units of $1,000).

Table 2: Full summary statistics in mTurk sample

	Mean	Std. Dev.	N
Age	33.04	11.09	2041
Female	0.44	0.50	2041
White	0.77	0.42	2041
Black	0.07	0.26	2041
Hisp	0.06	0.23	2041
Asian	0.08	0.27	2041
Income ($1,000)	49.47	39.33	2040
Fulltime	0.43	0.49	2041
Partime	0.14	0.35	2041
College	0.45	0.50	2041
Student	0.11	0.32	2041
Supported Obama in 2012	0.64	0.48	2040
R-loving (losses)	1.44	1.06	1534
R-loving (gains)	0.82	0.91	1572
Δ Wage unfairness for current v. new worker (cont)	1.41	1.02	1858
Higher WTA if bought house at $300K (binary)	0.79	0.41	1582

Notes: See Section 3.2 for a detailed description of the loss aversion variables (the last four variables in the Table). Briefly, "R-loving (losses)" takes integer values from $[0, 3]$, increasing in the number of times you choose the lottery option over the risk free option over options involving losses. "R-loving (gains)" is defined analogously, but over gains. "Δ wage unfairness" is increasing in how much more unfair a respondent deems a wage cut to a current versus a new employee. "Higher WTA" refers to measures of anchoring bias to the original sales price of a house.

Table 3: Further summary statistics and experimental balance

	(1) Cont. mean	(2) Tr. mean	(3) Diff.	(4) P-val
Age	33.26	32.94	0.326	0.513
Female	0.446	0.436	0.0105	0.637
White	0.779	0.769	0.00952	0.611
Black	0.0692	0.0790	-0.00977	0.404
Hisp	0.0536	0.0574	-0.00383	0.709
Asian	0.0770	0.0738	0.00315	0.790
Income ($1,000)	49.11	49.18	-0.0756	0.965
Fulltime	0.430	0.429	0.00111	0.960
Partime	0.149	0.129	0.0199	0.199
College	0.447	0.447	0.000189	0.993
Student	0.102	0.121	-0.0187	0.184
Supported Obama in 2012	0.643	0.636	0.00775	0.718
R-loving (losses)	1.458	1.428	0.0303	0.582
R-loving (gains)	0.802	0.846	-0.0438	0.349
Δ Wage unfairness for current v. new worker (cont)	1.361	1.450	-0.0894	0.0621
Higher WTA if bought house at $300K (binary)	0.792	0.794	-0.00237	0.908
Observations	1030	987	2017	2017

Notes: Observation totals are the shared non-missing observations across all variables. Col. (1) displays means for those randomized into the control version of the $5/$15 money-transfer experiment (where recipients do not know their original endowment) and col. (2) displays means for the treatment version (where recipients do know their original endowment). Col. (3) subtracts col. (2) from (1) and Col. (4) is the p-value associated with $H_0 : Diff = 0$.

Table 4: Effect of treatment on amount redistributed (main between-subject results)

	Dept. var: Amount redistributed							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated in	-0.797***	-0.827***	-0.829***	-0.923***	-0.824***	-0.763**	-0.669*	-0.509*
first stage	[0.104]	[0.106]	[0.134]	[0.103]	[0.134]	[0.312]	[0.369]	[0.304]
Cont. gp. mean	4.553	4.585	4.682	4.542	4.683	4.451	4.787	4.625
Controls	No	No	No	No	Yes	No	No	No
Ex. short duration	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ex. presented second	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Ex. incr. ineq	No	No	No	Yes	No	No	No	No
Ex. hypothetical	No	No	No	No	No	Yes	No	No
Ex. emphasis	No	No	No	No	No	No	Yes	No
Ex. promised payment	No	No	No	No	No	No	No	Yes
Observations	2044	1903	1227	1151	1220	195	183	191

Notes: *Treatment* indicates that the individual was considering two recipients who had *already been told their endowments*. Note that redistributing $5 results in complete equality between the two recipients. All regressions include session fixed effects. Controls: age, female, white, black, Hispanic, asian, income, student status, full-time status, part-time status, Obama support, 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. Ex. hypothetical: exclude survey sessions where the redistribution experiment was entirely hypothetical. Ex. emphasis: exclude survey sessions where the instructions to the redistribution experiment included underlined and italicized text and a reminder to the right of the person A and B bar chart. Ex. promised payment: include only the survey session where the instructions to the treatment arm of the redistribution experiment specify that the two other mTurk participants have been told the amount they initially received might increase or decrease. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 5: Effect of treatment on complete redistribution (between-subject extensive margin results)

	Dep. v.: Complete redistribution			
	(1)	(2)	(3)	(4)
Treated in first stage	-0.199*** [0.0250]	-0.203*** [0.0241]	-0.200*** [0.0250]	-0.211*** [0.0627]
Cont. gp. mean	0.820	0.869	0.821	0.824
Controls	No	No	Yes	No
Ex. incr. ineq	No	Yes	No	No
Ex. hypothetical	No	No	No	Yes
Observations	1227	1151	1220	195

Notes: All regressions include session fixed effects. Both subjects who finished the survey very quickly and subjects not presented the distribution experiment first were excluded from these regressions. $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$

Table 6: Effect of treatment on amount redistributed (within-subject results)

	Dept. var: Amount redistributed					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment stage	-0.762*** [0.0803]	-0.703*** [0.126]	-0.812*** [0.103]	-0.646*** [0.193]	-0.727*** [0.235]	-0.309* [0.181]
Cont. gp. mean	4.665	4.646	4.682	4.364	4.973	4.586
Sample	All	T→C	C→T	All	All	All
Ex. hypothetical	No	No	No	Yes	No	No
Ex. emphasis	No	No	No	No	Yes	No
Ex. promised payment	No	No	No	No	No	Yes
Observations	2373	1131	1242	390	366	382

Notes: All regressions include respondent fixed effects. Subjects who finished the survey very quickly and were not presented the distribution experiment first were excluded. $C \to T$ denotes the subsample that was first randomized into the control scenario and *then* the treatment scenario. $T \to C$ denotes the subsample that was first randomized into the treatment scenario and *then* the control scenario. $^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$

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

	Dep. v.: Amount redistributed		Dep. v.: Complete redistribution	
	(1)	(2)	(3)	(4)
Treated in first stage	-1.806***	-1.641***	-0.301***	-0.257***
	[0.360]	[0.368]	[0.0664]	[0.0682]
Cont. gp. mean	4.515	4.510	0.699	0.696
Controls	No	Yes	No	Yes
Observations	206	205	206	205

Notes: All regressions include session fixed effects. Subjects who finished the survey very quickly or were not presented the distribution experiment first were excluded from these regressions.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 8: Interacting loss aversion measures with treatment status

	Dept. var: Amount redistributed					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated in first stage	-0.829*** [0.134]	-0.130 [0.278]	-0.712*** [0.223]	-0.526*** [0.177]	-1.009*** [0.355]	-1.070*** [0.248]
Tr. x R-loving (losses)		-0.395** [0.156]				
Tr. x R-loving (gains)			0.0364 [0.181]			
Tr. x R-loving (L-G)				-0.237** [0.114]		
Tr. x Higher WTA					0.475 [0.400]	
Tr. x Δ Wage unfairness						0.221 [0.140]
Cont. gp. mean	4.682	4.671	4.649	4.641	4.671	4.695
Observations	1227	791	811	781	812	1055

Notes: All regressions include session fixed effects. "Tr. x R-loving (losses)", "Tr. x R-loving (gains)", and "Tr. x R-loving (L-G)" refer to the interaction of the the risk-loving over losses, risk-loving over gains, and risk-loving over losses *relative* to that over gains variables with $Treat$. "Tr. x Higher WTA" refers to the interaction of the variable that indicates that the respondent demanded a higher house price with a $300,000 initial price than the $250,000 price and $Treat$. "Tr. Δ Wage unfairness" refers to the interaction of the difference between the fairness ratings of the cut to the wages of the current and new coffee shop employees and $Treat$. "Inc. Covar x Treat" means that interactions with $Treat$ and the following list of variables are all included simultaneously: age, female, white, income, student status, full-time status, Obama support, and college degree. The main effects of these interactions are controlled for as well in columns 7 and 8. ${}^{*}p < 0.1, {}^{**}p < 0.05, {}^{***}p < 0.01$

Table 9: Preferred average tax rate for person who became rich five versus one year ago

	Dept. var: Chosen tax rate						Dept. var: Favor tax hike	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated (rich for five yrs.)	-0.0117* [0.00644]	-0.0171** [0.00751]	-0.0168*** [0.00630]	-0.0168*** [0.00631]	-0.0185*** [0.00709]	-0.0120 [0.0134]	-0.0735* [0.0415]	-0.0728* [0.0407]
Cont. gp. mean	0.288	0.286	0.288	0.288	0.290	0.284	0.777	0.777
Ex. if presented second	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. for outliers?	No	No	Yes	Yes	Yes	Yes	n.a.	n.a.
Controls	No	No	No	Yes	No	No	No	Yes
Vignette	Both	Both	Both	Both	Takeover	Lottery	Both	Both
Observations	1097	721	694	682	513	181	446	438

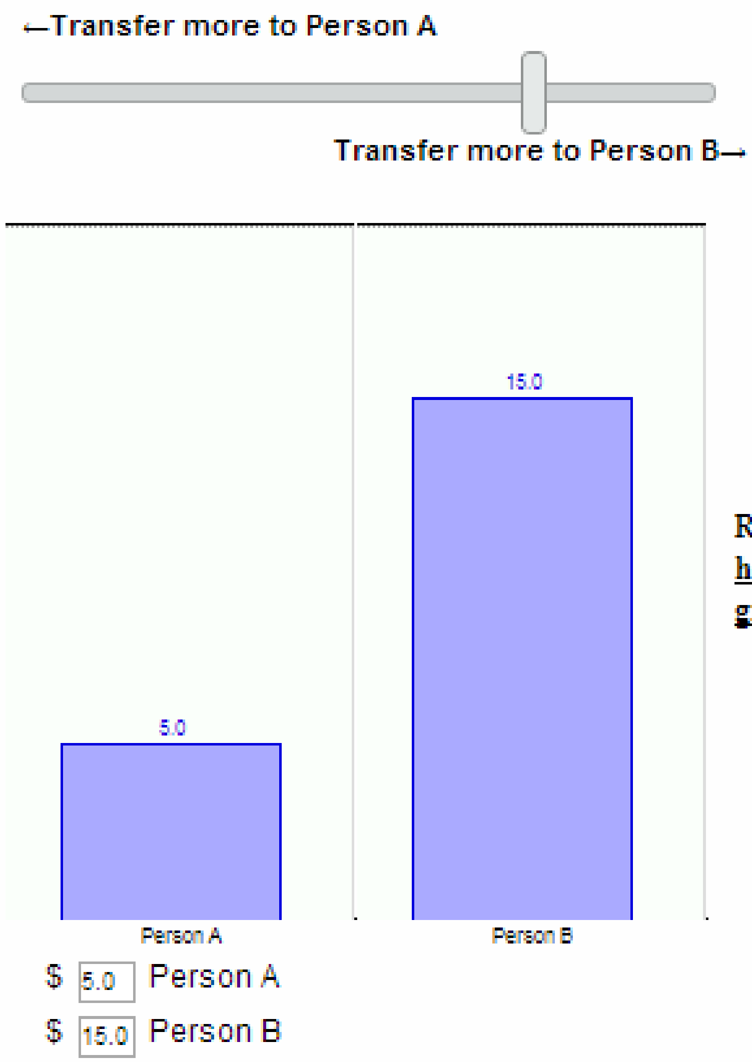
Notes: All regressions include session fixed effects. Subjects who finished the survey very quickly are excluded from the regression. "Adj. for outliers" indicates that the lowest five percent and the highest one percent of chosen tax rates are dropped (the asymmetry is due to a small mass of zeros, the default position of the slider). "Vignette" refers to the brief description of the event that led to the sudden increase in earnings. "Controls" refer to those specified in Table 4. Note that the outcomes in cols. (7) and (8) were not asked in all waves. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Appendix Figure 1: Main redistribution experiment (control arm)

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, they 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.



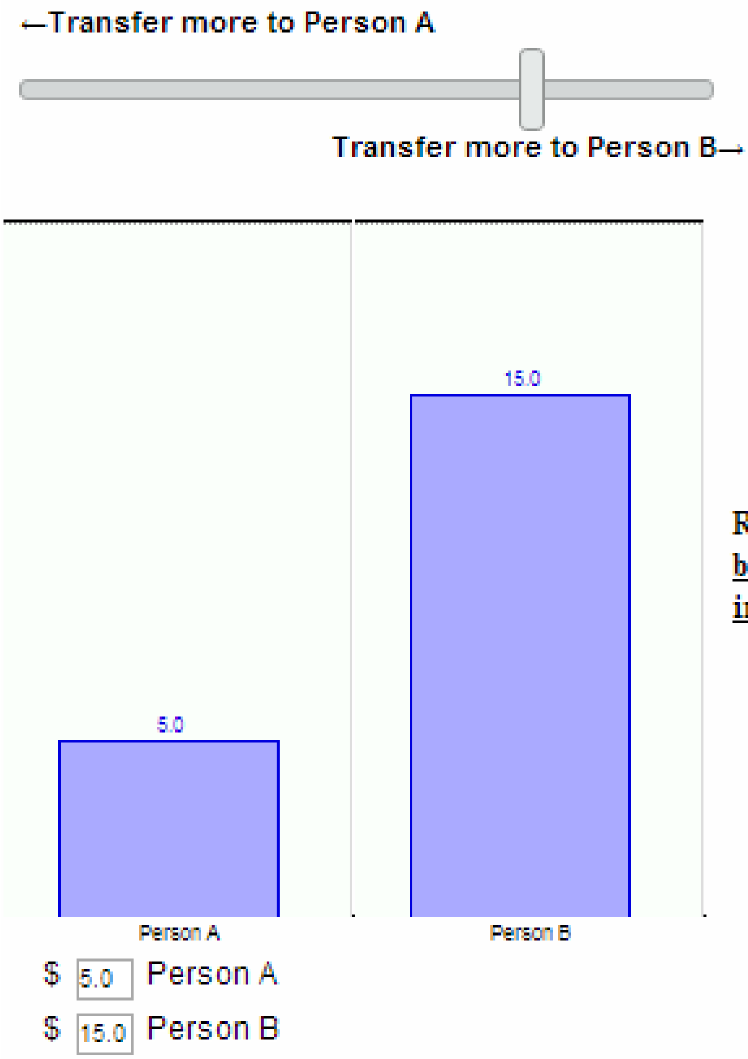
Recall: Person A and B do NOT know how much money they were initially given.

Appendix Figure 2: Main redistribution experiment (treatment arm)

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 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.

Please indicate below what transfer, if any, you would make.



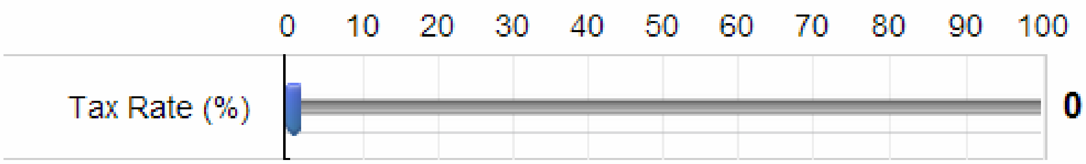
Recall: Person A and B have already been told how much money they were initially given.

Appendix Figure 3: 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. *<u>This year</u>*, 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.)

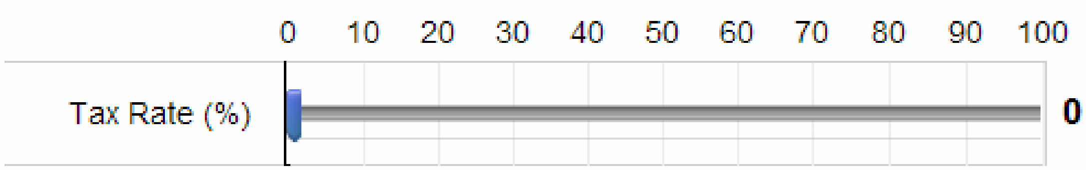


Appendix Figure 4: Income tax experiment (treatment arm)

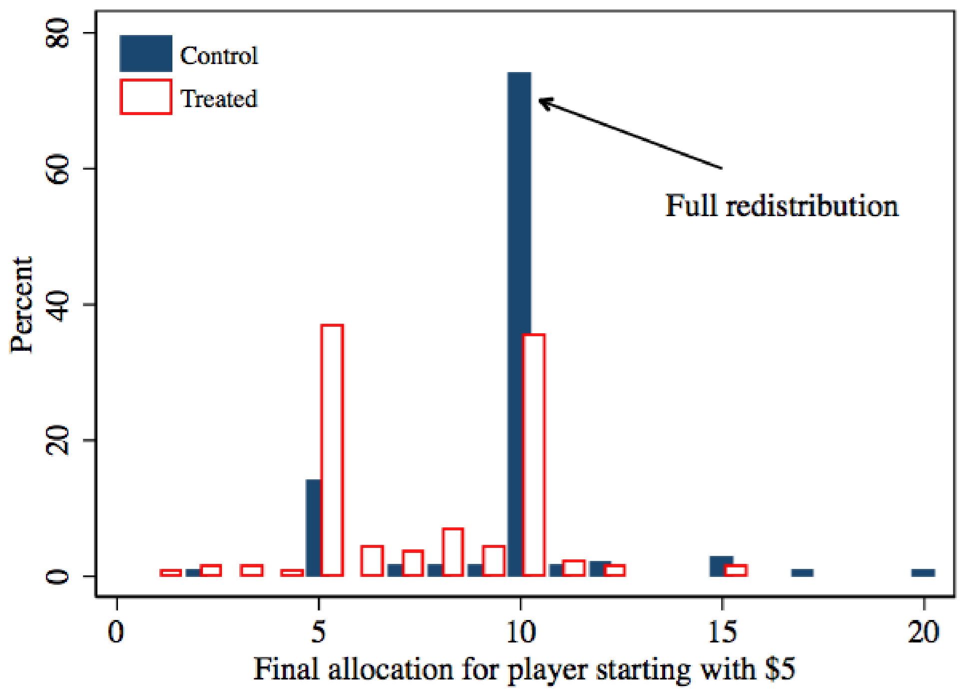
There has been much talk about whether wealthy families are paying their fair share in taxes.

Consider the following person. He started five years ago as a regional sales manager at a medium-sized firm. *<u>Soon after starting</u>*, 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.)



Appendix Figure 5: Histogram of ex-post allocations for the ex-ante poorer player, luck (control) versus merit (treatment) session



.

Appendix Table 1: Survey Session Details

Session	Date	Survey Obs	Analysis Obs	First Exper.	\$5/\$15 version	\$250,000 Tax Q
One	Feb 13, 2014	187	187	\$5/\$15	Standard	No
Two	Feb 27, 2014	312	295	\$5/\$15	Standard	No
Three	Mar 10, 2014	301	250	\$5/\$15	Standard	No
Four	Mar 21, 2014	352	282	Tax	Standard	Yes
Five	Mar 24, 2014	374	303	Tax	Standard	Yes
Six	Apr 25, 2014	321	228	\$5/\$15	Standard	No
Seven	May 28, 2014	312	207	\$5/\$15	Standard	Yes
Eight	May 30, 2014	332	216	\$5/\$15	Real Money	Yes
Nine	Jun 19, 2014	314	200	Tax	No Emphasis	Yes
Ten	Dec 11, 2014	307	196	\$5/\$15	No Promises	No
Recontact	March 2016	72	72	n.a.	n.a.	n.a.

Notes: Total observations and analysis sample observations differ because in almost all analysis we drop anyone who took a previous survey. Further details on each session can be found in the text. For the recontact survey, we recontacted respondents from Round 2 on March 22, 2016. They were provided with an invitation to an mTurk task that linked them to the re-contact survey. The re-contact survey asked only questions about own loss aversion and demographics (and thus does not include any experiment). We sent three reminders and kept the survey open for 36 days. Of the 301 respondents in session three, 72 took the re-contact survey, for a follow-up rate of 24 percent.

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	0.0241 [0.0168]	0.00182 [0.0106]	-0.0307 [0.0198]	0.553 [0.368]
Cont. gp. mean	0.0829	0.0341	0.878	11.43
Observations	1216	1216	1216	1057

Notes: Subjects who finished the survey very quickly and were not presented the distribution experiment first were excluded from these regressions.

Appendix Table 3: Predictive power of original responses for responses in follow-up survey (selected variables)

	(1) Female	(2) Obama 2012	(3) Gov't red.	(4) Loss av. (house)	(5) Loss av. (wage)
Original answer	1	0.917***	0.595***	-0.0473	0.190**
	[.]	[0.0507]	[0.0961]	[0.119]	[0.0915]
Observations	72	72	72	72	72

Notes: The sample is limited to the 25% of session three respondents who took-up our follow-up survey 24 months later. "Obama" refers to stated support for Barack Obama in the 2012 election. "Gov redist." is a 1-7 categorical variable measuring respondents support for government taking steps to shrink income disparities (taken from the General Social Survey). "Loss av. (house)" refers to our standard loss aversion question on willingness to accept a price for a house and "Loss av. (wage)" refers to our standard loss aversion question on assessing the fairness of wage cuts. Exact wording of each loss aversion question can be found later in the Appendix.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Appendix Table 4: Interacting loss aversion measures with treatment status (recontact sample)

	Dep. v.: Amount redistributed			
	(1)	(2)	(3)	(4)
Treated - First Stage	-0.846*	-2.000**	-0.0977	-1.305**
	[0.467]	[0.840]	[1.116]	[0.592]
Higher WTA		-0.524		
		[0.710]		
Treat x Higher WTA		1.627		
		[1.011]		
Δ Wage unfairness			-0.0821	
			[0.565]	
Treat x Δ Wage unfairness			-0.545	
			[0.709]	
Endowment effect				-0.0567
				[0.198]
Treat x Endowment effect				0.308
				[0.284]
Observations	72	72	72	71

Notes: Data are taken from the 25% of session three respondents to took-up the follow-up survey 24 months later. "Treated - first stage" indicates that the individual was in the treatment group *during the original session three survey* (there is no experiment in the recontact survey, just loss aversion and demographic questions). "Treat x Higher WTA" refers to the interaction of the variable that indicates that the respondent demanded a higher house price with a \$300,000 initial price than the \$250,000 price and $Treat$. "Treat x Δ Wage unfairness" refers to the interaction of the difference between the fairness ratings of the cut to the wages of the current and new coffee shop employees and $Treat$. "Treat x Endowment effect" refers to the interaction between the willingness to pay versus willingness to sell question and $Treat$. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table 5: Correlation across loss aversion measures

	R-loving (losses)	R-loving (gains)	R-loving (L-G)	Higher WTA	Δ Wage unfairness
R-loving (losses)	1				
R-loving (gains)	-0.0318	1			
R-loving (L-G)	0.765***	-0.668***	1		
Higher WTA	0.0937***	0.00588	0.0665**	1	
Δ Wage unfairness	-0.00906	0.0518**	-0.0461*	0.0794***	1
Observations	1871				

Notes: All of these responses were collected in the same survey as either the \$5/\$15 experiment or the tax experiment (i.e., were not part of the follow-up survey). The loss aversion questions are described in Section 3.2 and their exact wording is given later in the Appendix.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table 6: Differential treatment effects (main between-subject results)

	Dept. var: Amount redistributed											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated in first stage	-1.537***	-0.904***	-0.920***	-0.819***	-0.834***	-0.811***	-0.731***	-0.851***	-0.855***	-0.814***	-0.636***	-0.621***
	[0.428]	[0.181]	[0.291]	[0.139]	[0.139]	[0.139]	[0.222]	[0.141]	[0.179]	[0.146]	[0.182]	[0.226]
Tr. x Age	0.0211*											
	[0.0120]											
Tr. x Female		0.180										
		[0.271]										
Tr. x White			0.123									
			[0.328]									
Tr. x Black				-0.0623								
				[0.544]								
Tr. x Hisp					0.158							
					[0.555]							
Tr. x Asian						-0.180						
						[0.534]						
Tr. x Income							-0.00000194					
							[0.00000370]					
Tr. x Student								0.292				
								[0.460]				
Tr. x Fulltime									0.0719			
									[0.272]			
Tr. x Partime										-0.0622		
										[0.378]		
Tr. x College											-0.412	
											[0.271]	
Tr. x Obama												-0.314
												[0.281]
Cont. gp. mean	4.683	4.683	4.683	4.683	4.683	4.683	4.683	4.683	4.683	4.683	4.683	4.683
Observations	1220	1220	1220	1220	1220	1220	1220	1220	1220	1220	1220	1220

Notes: All regressions include session fixed effects and the controls listed in Table 4. Subjects who finished the survey very quickly and were not presented the distribution experiment first were excluded from these regressions. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Appendix Table 7: Main between-subject redistribution and tax results, includes subjects that participated in multiple batches

	Dept. var: Amount redistributed	Dept. var: 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. 3 of Table 4 and the second replicates the analysis from col. 2 of Table 9. "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). $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Appendix Table 8: 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 (orig. spec.)	Qreg	Winsorize	Drop 0s	Drop regr.
Observations	694	721	721	717	602

Notes: The first column replicates the preferred specification from Table 9 (col. 3), where outliers below the bottom fifth percent and above the higher one percentile are dropped. "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). $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Exact wording of loss aversion survey questions

In this section, we display the questions we developed to assess respondents' own loss aversion. All questions can also be seen in the links to the survey we provided, but are included here for convenience. We also describe how we use these questions to construct the actual variables we use to measure loss aversion.

Risk-aversion in gains versus losses

> Imagine that you face the following decisions. For each decision, please examine both options and indicate the one that you prefer.
>
> - Decision One:
>   - A sure gain of $5.
>   - 50% chance to gain $8 and 50% chance to gain nothing.
> - Decision Two:
>   - A sure gain of $5.
>   - 50% chance to gain $11 and 50% chance to gain nothing.
> - Decision Three:
>   - A sure gain of $5.
>   - 50% chance to gain $14 and 50% chance to gain nothing.

Subjects then faced the same gambles, but in the loss domain:

> - Decision One:
>   - A sure loss of $5.
>   - 50% chance to lose $8 and 50% chance to lose nothing.
> - Decision Two:
>   - A sure loss of $5.
>   - 50% chance to lose $11 and 50% chance to lose nothing.
> - Decision Three:
>   - A sure loss of $5.
>   - 50% chance to lose $14 and 50% chance to lose nothing.

Variable construction: Our measure of risk-seeking varies from zero to three, depending on the number of lotteries the subject chooses over the "sure" option. We calculate this 0-3 measure separately for losses and gains.

Willingness to sell as function of purchase price

> Suppose you bought a house for $250,000 a few years ago. The housing market in your neighborhood has since declined, and you have seen houses very similar to yours sell for $200,000, though some sell for a bit more and some sell for a bit less. You expect the current housing market conditions in your neighborhood to remain relatively stable. You are planning to relocate in the coming year for a new job. Someone is interested in buying your house. What is the least you would be willing to accept as a sale price?
>
> Suppose you bought a house for $300,000 a few years ago. The housing market in your neighborhood has since declined, and you have seen houses very similar to yours sell for $200,000, though some sell for a bit more and some sell for a bit less. You expect the current housing market conditions in your neighborhood to remain relatively stable. You are planning to relocate in the coming year for a new job. Someone is interested in buying your house. What is the least you would be willing to accept as a sale price?

Variable construction: We take the difference ($300,000 versus $250,000) as a measure of reference dependence. Due to outliers, we prefer a binary measure of whether the WTA was higher in the $300,000 vignette.

Consideration of reference points in labor market transactions

> A small coffee shop has one employee who has worked there for six months and earns $10 per hour. The shop continues to do fairly good business, though unemployment in the area has increased due to a factory closure nearby. As a result, other small restaurants have now hired reliable workers at $8 an hour to perform jobs similar to those done by the coffee shop employee. The owner of the coffee shop reduces the employee's wage to $8. The owner's actions were: ○ Completely fair ○ Acceptable ○ Unfair ○ Very unfair.
>
> A small coffee shop has one employee who has worked there for six months and earns $10 per hour. The shop continues to do fairly good business, though unemployment in the area has increased due to a factory closure nearby. As a result, other small restaurants have now hired reliable workers at $8 an hour to perform jobs similar to those done by the coffee shop employee. The current employee leaves, and the owner decides to pay a replacement worker $8 an hour. The owner's actions were ○ Completely fair ○ Acceptable ○ Unfair ○ Very unfair.

Variable construction: We take the difference in fairness assessment (current versus new worker) as a measure of reference dependence. As each assessment varies from 1 to 4, the difference can vary from 3 to -3.

Willingness to pay versus willingness to accept the same object

Note that this question was asked only in the one-year follow-up survey.

> Suppose you were offered the chance to buy a lottery ticket with a 50% chance of paying \$0 and a 50% chance of paying \$10. At what price would you be willing to buy this lottery ticket?
>
> Suppose you owned a lottery ticket with a 50% chance of paying \$0 and a 50% chance of paying \$10. At what price would you be willing to sell this lottery ticket?

Variable construction: We take the difference (willingness to sell minus willingness to buy) as a measure of reference dependence.

Details from the excluded experimental session

In one session (April 17, 2014), instead of having individuals transfer money between the recipients using the slider, we tried to use a drop-down menu (which listed all twenty possible money transfers). In all cases, "no transfer" was listed first, which likely caused strong anchoring effects. More seriously, we only realized ex post that the drop-down menu covered up the graphic of the two recipients' endowments (which in the slider version respondents could see change in real time as they moved from one allocation to the other).

We report results from this session below but do not include them in our main results.

	Dept. var: Amount redistributed		
	(1)	(2)	(3)
Treated in first stage	-0.151	-0.467	-0.521
	[0.336]	[0.318]	[0.338]
Cont. gp. mean	3.747	3.713	3.699
Controls	No	No	Yes
Ex. short duration	Yes	Yes	Yes
Ex. incr. ineq	No	Yes	Yes
Observations	209	200	197

Notes: Subjects who finished the survey very quickly were excluded.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$