Who Should Get Money? Estimating Welfare Weights in the U.S.*

Francesco Capozza[†]
Krishna Srinivasan[‡]

November 1, 2023

[Most recent version]

Abstract

Evaluating the desirability of a reform commonly involves weighing the gains of the winners against the losses of the losers using welfare weights. Welfare weights measure the relative social value of providing individuals with an additional dollar of consumption. These weights can capture various normative ideals, such as utilitarianism or libertarianism. We elicit the welfare weights assigned by the general population of the U.S. using a real-stakes experiment. The general population weights can be used to identify socially acceptable policies. We find that the general population weights are more progressive than the weights implied by tax and transfer policies in the U.S., indicating that the general population desires additional redistribution. The general population weights are less progressive than the weights frequently used in the optimal policy literature. We explore the implications of these weights for optimal income taxes.

Keywords: Welfare Weights, Policy Views, Income Taxation
JEL Classification: C93, D31, H23, I31

---

†WZB and Berlin School of Economics. Email: francesco.capozza@wzb.eu
‡University of Zurich. Email: krishna.srinivasan@econ.uzh.ch.

1 Introduction

Policy reforms typically result in winners and losers. Consider a tax reform that increases taxes on high-income individuals and redistributes these taxes to low and middle-income individuals. Evaluating the desirability of such a reform involves asking: How should society weigh the gains of the winners against the losses of the losers? This question is at the forefront of the policy debates in domains such as taxation, in-kind transfers, and social insurance. The literature in public economics commonly weighs the gains and losses accrued from a reform using welfare weights. Welfare weights measure the relative social value of providing individuals with an additional dollar of consumption. These welfare weights can capture various normative ideals, such as equality of opportunity or utilitarianism (Saez and Stantcheva, 2016). An open question in the literature is: Which welfare weights should society use to evaluate reforms?[1]

The literature makes different assumptions about individuals' utilities, which leads to a wide range of estimates for welfare weights. However, it is unclear which of these estimates would lead to "socially acceptable" policies. An advantage of basing policy recommendations on socially acceptable policies—those without widely held normative views against them—is that these policies are more likely to be implemented (Diamond and Saez, 2011).[2] In this paper, we elicit the welfare weights assigned by the general population of the U.S. These general population weights can be used to identify socially acceptable policies.

To elicit the welfare weights of the general population of the U.S., we conducted two real-stakes online experiments with large samples of the general population in the U.S. ($N \approx 4000$). In the experiments, participants in the role of "Social Architects" face real "Recipients." The Recipients' disposable incomes span the income distribution of the U.S. A Social Architect chooses between various monetary transfers. Their choices are used to identify the welfare weights they assign to the Recipients. Our controlled experimental approach allows us to elicit welfare weights unconfounded by views about the government and taxation.

In our framework, a reform is conditioned solely on Recipients' incomes. In the experiment, Social Architects assign welfare weights based on Recipients' incomes. Thus, knowing a Social Architect's welfare weights in the experiment is sufficient to learn their

---

[1]As documented in the literature, this question is important because it has implications for the progressivity of policies (e.g., Saez, 2002), government spending (Hendren and Sprung-Keyser, 2020), inequality (e.g., Hendren, 2020), and growth (e.g., Hendren, 2020).

[2]There is growing interest in economics to identify socially acceptable policies (e.g., Saez and Stantcheva, 2016; Weinzierl, 2014)

assessment of the welfare implications of a reform and, by extension, its desirability. Although a Social Architect's welfare weights can be guided by various underlying normative ideals, such as equality of opportunity or poverty alleviation, our approach allows for policy evaluation without the need to specify and uncover the underlying ideals. This "sufficient statistics" approach is powerful because uncovering underlying ideals can be challenging when Social Architects are guided by a wide array of normative ideals.

We find that the welfare weights across all Social Architects are "progressive" on average, meaning that they are decreasing with Recipients' incomes. Since the welfare weights were elicited given the Recipients' disposable income resulting from the current tax and transfer system, our finding indicates that, on average, the general population wants additional redistribution beyond that achieved by the current tax and transfer system. To estimate the progressivity of the welfare weights, we estimate the elasticity of Social Architects' welfare weights with respect to Recipients' incomes. Our elasticity estimate implies that if a Social Architects assign a weight of 1 to a Recipient, they would assign a weight of 66 cents to a Recipient earning twice as much. This means that for every taken from a Recipient, Social Architects are willing to give as little as 66 cents to a Recipient earning half as much. We compare our elasticity estimate to the benchmark categories proposed in the literature (e.g., Allcott et al., 2019; Saez, 2002). This comparison indicates that the general population weights are characterized by "weak" to "strong" redistributive tastes.

While the general population weights are progressive on average, there is considerable heterogeneity in the progressivity of the weights. We find that 66% of the Social Architects have progressive weights, weights decreasing with Recipients' incomes, while 34% of the Social Architects have regressive weights, weights increasing with Recipients' incomes. A share of these Social Architects with regressive weights may be libertarians, which means that they do not want any redistribution. Our results align with findings from Almås et al. (2020), who find, in a different setting, that 30% of the participants sampled from a nationally representative population of the U.S. are libertarians.

Is the observed heterogeneity in welfare weights due to heterogeneous underlying normative ideals? We present two results to support the hypothesis that Social Architects' welfare weights are driven by their underlying ideals. First, we included two questions that are used to test whether Social Architects' welfare weights capture "non-welfarist" ideals in addition to the traditional "welfarist" ideals. In the welfarist approach, welfare weights depend on characteristics affecting Recipients (e.g., disability status). In the non-welfarist approach, welfare weights depend on characteristics not affecting Recipients (e.g., parental income). The non-welfarist approach includes a wide range of ideals, such

as redistribution based on the source of income. We find that Social Architects' welfare weights capture welfarist and non-welfarist ideals. Second, we include a question that is designed to capture an important non-welfarist ideal documented in the literature: redistribution based on the source of income. We find that Social Architects' responses to this question predict their welfare weights, implying that their welfare weights capture this ideal. These two results provide suggestive evidence that the heterogeneity in welfare weight is likely driven by the heterogeneity in underlying ideals.

The heterogeneity in the general population weights is partly due to their background characteristics. Republicans assign less progressive welfare weights relative to Democrats and Independents. This result suggests that a part of the partisan gap in support for redistribution documented in the literature (Fisman et al., 2017; Singhal, 2008) is likely driven by the partisan gap in welfare weights. Social Architects with higher incomes assign less progressive weights. This result suggests that a part of the gap in support for redistribution between low and high-income individuals documented in the literature (Cohn et al., 2019; Singhal, 2008) is likely due to the gap in welfare weights.

Next, we explore the empirical link between Social Architects' welfare weights and their support for government redistribution. This exercise highlights the value of using the general population weights to identify socially acceptable policies. Our results suggest that Social Architects' assigned welfare weights predict their support for government redistribution. A benchmarking exercise reveals that Social Architects' support for redistribution can be predicted with similar accuracy using either their assigned welfare weights or their stated political affiliation. These results suggest that the general population weights can be used to evaluate optimal policies to ensure their social acceptability.

How do the general population weights compare to the weights implied by the income tax schedule and transfer policies in the U.S.? The weights implied by the tax schedule and transfer policies can be represented as politicians' aggregation of societal welfare weights, potentially influenced by political economy considerations. We obtain the weights implied by the tax schedule from Hendren (2020) and transfer policies from Hendren and Sprung-Keyser (2020). Our elasticity estimates imply that the general population weights are 3.7 to 5.3 times more progressive than the weights implied by the tax schedule and 1.4 to 2 times more progressive than the weights implied by transfer policies. Representing the weights implied by the tax schedule as politicians' aggregation of societal welfare weights, we explore the factors related to aggregation that can explain the gap between the general population weights and the weights implied by the tax schedule. Our results are consistent with the explanation that politicians, aggregating societal welfare weights, overweight high-income individuals who assign less progressive weights.

This explanation is consistent with the evidence in the literature suggesting that implemented policies are often more likely to reflect the interests of high-income individuals (e.g., Gilens and Page, 2014; Karabarbounis, 2011).

How do the general population weights compare to the weights used in the optimal policy literature? One frequently used estimate of welfare weights is inversely proportional to Recipients' disposable incomes or consumption (e.g., Saez, 2001), which we refer to as "inverse-consumption" weights. These weights can be obtained by assuming log utilities or constant relative risk aversion (CRRA) utilities in consumption with a coefficient of relative risk aversion of 1. We find that inverse-consumption weights are 1.8 to 3.3 times more progressive than the general population weights. Our results suggest that the literature could use CRRA utilities, but with a coefficient of relative risk aversion ranging from 0.37 to 0.53, to obtain policies that would be socially acceptable in the U.S.

We explore the implications of the estimated welfare weights by calibrating the optimal non-linear labor income taxes. We find that the optimal marginal tax rates calibrated with the general population weights are lower than the rates calibrated with inverse-consumption weights but higher than the current income tax schedule.

Our paper is related to four strands of literature. The first strand of literature aims to directly elicit people's welfare weights using experiments. The closest paper to ours is by Saez and Stantcheva (2016). In their experiment, participants recruited from the online labor market Amazon Mechanical Turk face pairs of hypothetical families with different disposable incomes and taxes and are asked which family in each pair is more deserving of a $1,000 tax break. Their paper calibrates the optimal linear income taxes with no behavioral responses, while our paper calibrates the optimal non-linear income taxes with behavioral responses. There are two other important differences. First, our paper uses samples that are broadly representative of the general population, while their paper uses a non-representative sample. Second, their paper uses hypothetical decisions, while our paper uses decisions with real stakes. Moreover, our paper explores whether the general population weights can be used to obtain socially acceptable policies, explores the predictors of the weights, and compares these weights to the weights implied by tax and transfer policies in the U.S. and the weights frequently used in the optimal policy literature.

The second strand of literature identifies the welfare weights implied by the tax schedule (e.g., Bourguignon and Spadaro, 2012; Hendren, 2020; Lockwood and Weinzierl, 2016; Zoutman et al., 2013) or by transfer policies (Hendren and Sprung-Keyser, 2020). These "inverse-optimum weights" represent politicians' (implied) aggregation of societal welfare weights, potentially influenced by their political economy considerations. There are

several limitations to using these weights to evaluate other policies (see Lockwood and Weinzierl (2016); Stantcheva (2016)). First, they may not reflect societal preferences if politicians are influenced by political economy considerations, such as lobbying. Second, they can sometimes be negative, in which case, they cannot be used in standard policy formulas that require positive welfare weights. Third, they are sensitive to the assumptions about the elasticity of taxable income.

The third strand of literature aims to incorporate normative ideals in optimal policy formulas by modifying individuals' utilities or the objective function (e.g., Fleurbaey and Maniquet, 2006; Weinzierl, 2014, 2018). A key limitation of this approach is the difficulty in incorporating multiple ideals. In our paper, we elicit the welfare weights of the general population, which can capture various underlying ideals, and use these weights to evaluate "standard" optimal policy formulas. Our approach is supported by our results suggesting that Social Architects' welfare weights are heterogeneous, likely guided by heterogeneous underlying ideals.

The fourth is the experimental literature that uses controlled experiments to identify the ideals that guide people's welfare weights and, consequently, their support for redistribution (e.g., Almås et al., 2020; Drenik and Perez-Truglia, 2018). These papers do not directly elicit welfare weights. In our paper, we elicit the welfare weights of the general population of the U.S., which can be guided by various underlying ideals.

The structure of our paper is as follows. In Section 2, we present the theoretical framework based on the framework of Saez and Stantcheva (2016). This section provides a mapping from the theory of optimal taxation to the experimental design. Section 3 presents the experimental design. Section 4 presents the results on the general population weights. Section 5 compares the general population weights to the weights implied by tax and transfer policies and the weights used in the optimal policy literature. Finally, Section 6 discusses and concludes.

2   Theoretical Framework

Our theoretical framework is based on the framework by Saez and Stantcheva (2016). It describes the mapping from the theory of optimal taxation to the experimental design described in the next section.

2.1   Recipients' Utilities

Consider a population of N *Recipients* indexed by j. A Recipient j's indirect utility function is given by $U_j = z_j - T_j(z_j) - v(z_j, \alpha_j)$ where z_j denotes Recipient j's optimal income (implicitly based on their optimal choice of labor) given the tax schedule, T_j denotes the taxes paid by Recipient j given their optimal income, and v represents the disutility of

work.[3] The disutility of work increases with income. Implicitly, this captures the idea that achieving higher income requires higher effort. The disutility of work can also depend on various personal characteristics α_j (e.g., disability status).

2.2 Welfare Weights

A *Social Architect* assigns *generalized social marginal welfare weights* (henceforth *welfare weights*) to the Recipients. The welfare weight g_j measures the relative social value of a $1 increase in consumption to Recipient j (relative to other Recipients) and is given by $g_j = g(c_j, \theta_j)$. In this equation, $c_j = z_j - T_j(z_j)$ is the consumption of Recipient j and the vector θ_j includes a set of Recipients' characteristics.[4] Some characteristics in θ_j may also be included in α_j (e.g., disability status), while others may not (e.g., parental income). The welfare weights are relative and thus defined up to a multiplicative constant.

A Social Architect's welfare weights can be guided by various ideals, such as equality of opportunity, utilitarianism, distribution based on the source of income, and poverty alleviation.[5] For example, a utilitarian Social Architect would assign welfare weights proportional to Recipients' marginal utility of consumption (captured by c_j). A Social Architect guided by equality of opportunity would assign higher welfare weights to Recipients from disadvantaged backgrounds (captured by θ_j) than those from more advantageous backgrounds.

2.3 Evaluating Reforms

Consider a setting with one Social Architect and two Recipients (selected from the N Recipients). The Recipients have incomes z_l and z_h, such that $z_h > z_l$. We consider a "small" (marginal) budget-neutral reform. The reform $\epsilon R = (\epsilon r_l, -\epsilon r_h)$ with $r_h = r_l$ changes the Recipients' disposable incomes by taking the monetary amount ϵr_h from the higher-income Recipient and giving the amount ϵr_l to the lower-income Recipient.[6]

A Social Architect may assign different welfare weights to Recipients earning the same income. Since the reform is conditioned only on Recipients' incomes, the welfare weights, which are at a more disaggregated level, have to be aggregated up to the level of the reform. We assume that a Social Architect reports their mean welfare weights.[7] The mean

---

[3]There is no common concave transformation of the utilities, implying that all the utility gains and losses are expressed in dollar terms. The welfare weights estimation does not rely on this assumption.

[4]In the empirical exercise of Saez and Stantcheva (2016), the welfare weights also depend on the taxes (T_j). We assume that T_j is contained in θ_j.

[5]See Saez and Stantcheva (2016) for an overview of the ideals that can be incorporated by this approach.

[6]Since the Recipients are optimizing, the marginal reform does not affect their choice of labor and, consequently, their optimal pre-tax income z due to the envelope condition.

[7]When reforms are conditioned jointly on incomes and other observable characteristics ("tags"), there are two ways to evaluate reforms. The first approach involves averaging the weights up to the level of the reform, i.e., jointly on incomes and other observable characteristics. The second approach is to ignore

welfare weights $\bar{g}_j$ assigned to Recipient j earning z is the mean of the welfare weights assigned to all Recipients earning z. It is given by

$$\bar{g}_j = \frac{\sum_{j:z_j=z} g_j}{h(z)}. \tag{1}$$

where $h(z)$ is the number of Recipients with earnings z. A set of mean weights assigned to the Recipients can be consistent with multiple underlying ideals. For example, utilitarianism and equality of opportunity may lead to the same mean weights. A utilitarian Social Architect would assign welfare weights proportional to Recipients' marginal utility of consumption. Since the marginal utility of consumption decreases with income, mean welfare weights also decrease with income. A Social Architect guided by equality of opportunity would assign higher welfare weights to Recipients from disadvantaged backgrounds. Since the share of Recipients from disadvantaged backgrounds decreases with income, mean welfare weights also decrease with income. There exists a distribution of Recipients from disadvantaged backgrounds that makes the mean welfare weights corresponding to these two ideals identical.

A Social Architect maximizes a utilitarian social welfare function. The social welfare function is the weighted sum of utilities, weighted by the mean welfare weights, and is given by

$$\begin{aligned} W &= \bar{g}_l \cdot U_l + \bar{g}_h \cdot U_h \\ &= \bar{g}_l \cdot (z_l - T_l(z_l) + \epsilon r_l - v(z_l, \alpha_l)) + \bar{g}_h \cdot (z_h - T_h(z_h) - \epsilon r_h - v(z_h, \alpha_h)). \end{aligned} \tag{2}$$

Taking the first derivative of Equation (2) with respect to ϵ, we get

$$\Delta W = \bar{g}_l \cdot r_l + \bar{g}_h \cdot (-r_h). \tag{3}$$

Equation (3) shows that the weighted reform amounts, weighted by the mean welfare weights, has a first-order effect on Recipients' welfare. The reform $R = (r_l, -r_h)$ is defined as being desirable if $\Delta W > 0$. Saez and Stantcheva (2016) derive the necessary conditions for a tax system to be at a local optimum: If a tax system is at a local optimum, then for

---

the tags, compare the reforms based on their impact on Recipients with various incomes, and average the weights up to Recipients' incomes. The latter approach is described in Hendren and Sprung-Keyser (2020).

any small budget-neutral reform R, $\Delta W = 0$.

2.4 Identifying Welfare Weights

To identify the mean welfare weights $\bar{g}_l$ and $\bar{g}_h$, we identify a non-budget-neutral reform $R' = (r_l, -r_h)$ with $r_l \neq r_h$ that makes the Social Architect indifferent between the reform R' and the budget-neutral reform $R = (r_l, -r_h)$ with $r_l = r_h$. Setting $\Delta W(R) = \Delta W(R')$ allows to recover the welfare weights $\bar{g}_l$ and $\bar{g}_h$.[8]

In our framework, a reform is conditioned solely on Recipients' incomes. Thus, knowing a Social Architect's mean welfare weights is sufficient to learn their assessment of the welfare implications of a reform (ΔW) and, by extension, its desirability. Even though a set of mean welfare weights may be consistent with multiple underlying ideals, these ideals have the same welfare implications. For example, a Social Architect may be guided by utilitarianism or equality of opportunity and may have the same mean weights in both cases. However, uncovering the exact ideal that guides the Social Architect is not important insofar as knowing their mean weights is sufficient to know their assessment of the welfare implications of the reform. This sufficient statistics approach is powerful because uncovering underlying ideals can be challenging when Social Architects are guided by a wide array of normative ideals.

3 Experimental Design

3.1 Eliciting Welfare Weights

Decisions

Participants in our experiment assume the role of either a "Social Architect" or a "Recipient." Each Social Architect faces a pair of real Recipients and learns about their real-world disposable incomes. We aim to identify the welfare weights the Social Architect assign based on Recipients' disposable incomes accrued from the current tax and transfer system. The Social Architect is informed that the Recipients receive a $1500 endowment. Next, the Social Architect makes several choices between monetary amounts that resemble policy reforms. These amounts are added to or subtracted from the Recipients' endowment. A Social Architect's choices are used to identify the welfare weights assigned to the Recipients. The Recipients are passive subjects who receive money based on the Social Architect's decisions.

As indicated in Section 2.3, to identify the welfare weights assigned by a Social Architect, we need to identify a non-budget neutral reform R' that makes a Social Architect in-

---

[8]Another approach is to identify the reform R' that makes the Social Architect indifferent between R' and the status-quo $R = (0,0)$.

different between R' and a budget-neutral reform R. A Social Architect is asked to choose between a "Balanced Reform" $R = (\$500, -\$500)$ and various "Unbalanced Reforms" of the type $R' = (\$rt, -\$t)$; the reforms $(\$rt, -\$t)$ take $\$t$ from the higher-income Recipient and give $\$rt$ to the lower-income Recipient. Our goal is to identify the reform $(\$rt, -\$t)$ that makes a Social Architect indifferent between $(\$rt, -\$t)$ and $(\$500, -\$500)$.[9] Setting $\Delta W((\$500, -\$500)) = \Delta W((\$rt, -\$t))$ in Equation (3), we get

$$\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l} = \frac{rt - 500}{t - 500}. \tag{4}$$

Equation (4) shows that $\tilde{g}$, the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, is an increasing function of the parameter r. When $r < 1$, $\tilde{g} < 1$, implying that the welfare weight assigned to the higher-income Recipient is lower than that assigned to the lower-income Recipient, i.e., $\bar{g}_h < \bar{g}_l$. Thus, $r < 1$ corresponds to "progressive" welfare weights. Intuitively, a Social Architect is willing to take a dollar from the higher-income Recipient even if this results in less than a \$1 benefit for the lower-income Recipient. Analogously, $r > 1$, $\tilde{g} > 1$ corresponds to "regressive" weights. Intuitively, a Social Architect is only willing to take a dollar from the high-income Recipient if this results in more than a \$1 benefit for the lower-income Recipient. Finally, $r = 1$, $\tilde{g} = 1$ corresponds to equal weights assigned to both Recipients.

Table 1 presents fifteen Unbalanced Reforms of the type $(\$rt, -\$t)$ used to elicit welfare weights. In Row 8, the Unbalanced Reform $(\$1000, -\$1000)$ corresponds to $r = 1$. Unbalanced Reforms above Row 8 correspond to $r < 1$, while those below Row 8 correspond to $r > 1$. Lower row numbers include Unbalanced Reforms with smaller values of r, which are, consequently, more progressive reforms. The maximum amount that can be taken from the higher-income Recipient corresponds to the initial endowment of \$1500. The minimum possible amount that can be taken from the higher-income Recipient corresponds to the \$500 in the Balanced Reform. The amount given to the lower-income Recipient is chosen so that the sum of the absolute value of the reform amounts for both Recipients is \$2000.

Unbalanced Reforms $(\$rt, -\$t)$ with larger values of r are less progressive. However, they are more efficient since they allocate larger amounts to lower-income Recipients while taking smaller amounts from higher-income Recipients. Our setting is characteristic of real-world settings in which the efficiency of policies comes at the cost of equity. It

---

[9]In principle, we can identify the reform $(\$rt, -\$t)$ that makes a Social Architect indifferent between $(\$rt, -\$t)$ and $(\$0, \$0)$. However, to minimize the Social Architect's susceptibility to status-quo bias, which entails choosing the status-quo $(\$0, \$0)$, we avoided this approach.

Table 1: Set of Reforms

Row	Balanced Reform	Unbalanced Reform	r
1	($500, −$500)	($550, −$1450)	0.38
2	($500, −$500)	($625, −$1375)	0.45
3	($500, −$500)	($700, −$1300)	0.54
4	($500, −$500)	($750, −$1250)	0.60
5	($500, −$500)	($800, −$1200)	0.67
6	($500, −$500)	($875, −$1125)	0.78
7	($500, −$500)	($950, −$1050)	0.90
8	($500, −$500)	($1000, −$1000)	1.00
9	($500, −$500)	($1050, −$950)	1.11
10	($500, −$500)	($1125, −$875)	1.29
11	($500, −$500)	($1200, −$800)	1.50
12	($500, −$500)	($1250, −$750)	1.67
13	($500, −$500)	($1300, −$700)	1.86
14	($500, −$500)	($1375, −$625)	2.20
15	($500, −$500)	($1450, −$550)	2.64

Notes: The table presents fifteen Unbalanced Reforms. An Unbalanced Reform ($rt, −$t$) takes t from the higher-income Recipient in the pair and gives rt to the lower-income Recipient.

is unavoidable to vary the progressivity of the reforms while keeping the efficiency of the reforms constant. Eliciting the progressivity of the welfare weights is akin to asking how much efficiency losses the Social Architect is willing to accept.

Our theory predicts that a Social Architect would choose a Balanced Reform starting from Row 1 in Table 1 before switching to an Unbalanced Reform. We identify the Unbalanced Reform that makes them indifferent between this reform and a Balanced Reform by calculating the mid-point of the Unbalanced Reform that a Social Architect switches to and the Unbalanced Reform in the previous row. If a Social Architect switches in the first (last) row, we calculate the mid-point of the Unbalanced Reform in the first (last) row and the bound on possible reforms.[10] Plugging the mid-point of the Unbalanced Reforms into Equation (4) allows us to estimate $\tilde{g}$.[11]

---

[10]If a Social Architect always chooses the Unbalanced Reform (switches in the first row), we take the mid-point of ($550, −$1450) and ($500, −$1500). The reform ($500, −$1500) takes the maximum possible amount from the higher-income Recipient, corresponding to their entire initial endowment. The reform ($500, −$1500) corresponds to $\tilde{g} = 0$. If a Social Architect always chooses the Balanced Reform (never switches), we take the mid-point of ($1450, −$550) and ($1500, −$500). The reform ($1500, −$500) takes the minimum possible amount from the higher-income Recipient, corresponding to the amount in the Balanced Reform. The reform ($1500, −$500) corresponds to an undefined $\tilde{g}$.

[11]For example, if a Social Architect prefers a Balanced Reform in Rows 1-6 and switches to the Unbal-
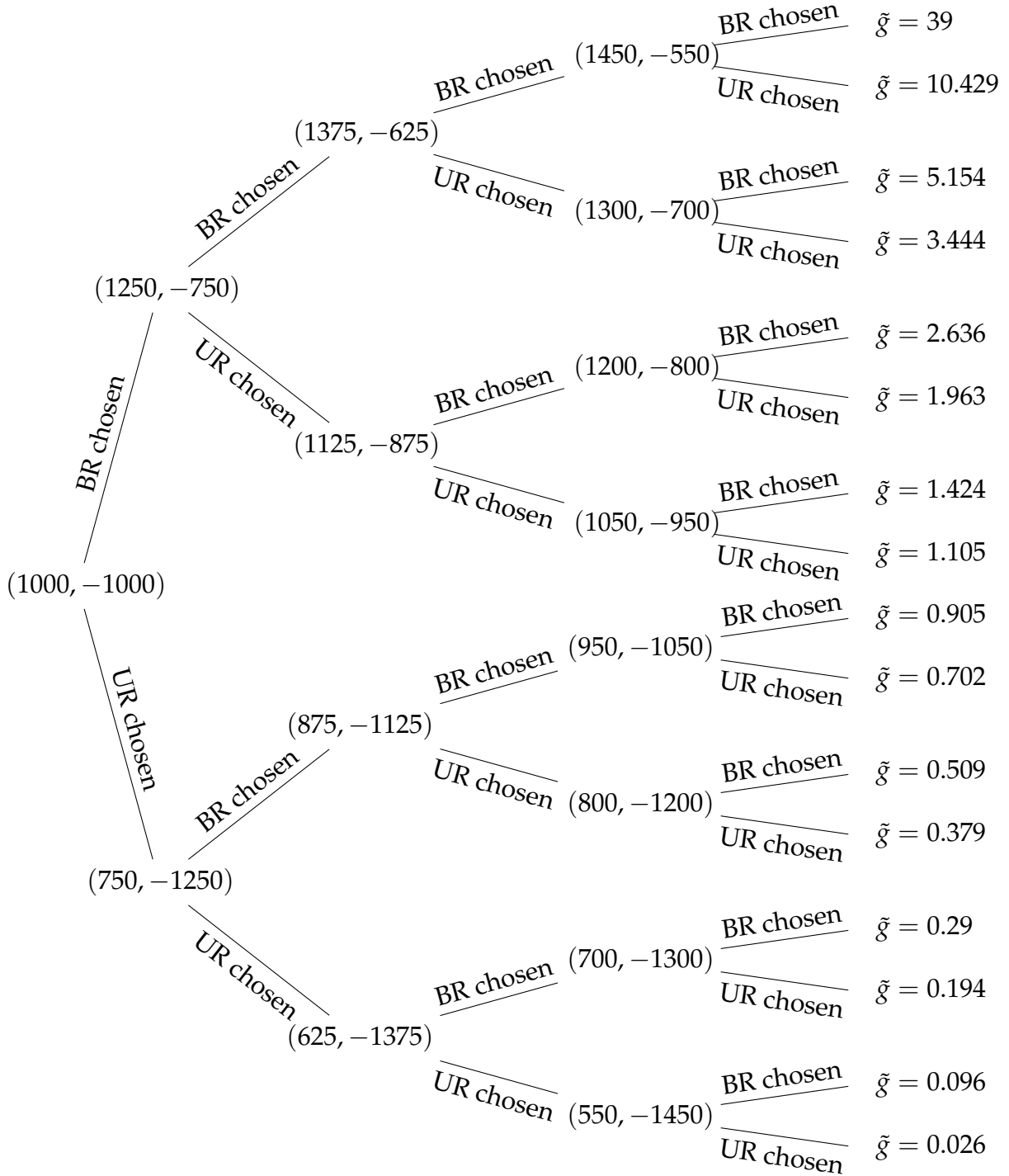
Figure 1: Unbalanced Reforms Selected by the Staircase Method

Notes: "UR chosen" and "BR chosen" indicate that the Unbalanced Reform and the Balanced Reform were chosen in the previous node, respectively. The Balanced Reform is $(500, -500)$. The parameter $\tilde{g}$ is the ratio of the weight assigned to the higher-income Recipient to the weight assigned to the lower-income Recipient.

To identify a Social Architect's switch-point in Table 1, we use the "staircase method" and present them with four decisions. The first decision, indicated in Row 8, asks a Social Architect to choose between the reforms ($500, −$500) and ($1000, −$1000). The second, third, and fourth decisions are selected adaptively, i.e., they depend on the choices of the Social Architect in the first, second, and third decisions, respectively.[12] Figure 1 presents a graphical representation of the selection of the four decisions. The staircase method enforces that a Social Architect has a unique switch-point and can only switch from a Balanced Reform to an Unbalanced Reform. It also ensures that a Social Architect is presented with decisions from only the top half of the table if they select a (progressive) Unbalanced Reform in the first decision and only the bottom half if they select a (regressive) Balanced Reform in the first decision. Thus, a Social Architect's response to the first decision determines if their welfare weights are progressive or regressive. This feature ensures that if a Social Architect's choice in the first decision is free from measurement error, the qualitative assessment of their welfare weights (whether progressive or regressive) is free from measurement error.

Decision Screens

Social Architects face a pair of Recipients in each "decision screen." Table 2 displays the incomes of the Recipients in each decision screen. For half the Social Architects, the order of the decision screens is reversed. The Recipients' incomes span the income distribution of the U.S., covering the various tax brackets. Figure 2 plots the incomes of the seven Recipients (dots) against the disposable income distribution (line) in the U.S. in 2019.

Table 2: Pairs of Recipients in Decision Screens

	Decision Screen					
	1	2	3	4	5	6
Recipient l	$8,000	$35,000	$70,000	$70,000	$70,000	$70,000
Recipient h	$70,000	$70,000	$100,000	$170,000	$250,000	$500,000

A Social Architect makes four choices in each decision screen. Their choices are used to identify $\tilde{g}$, the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, in each decision screen. Since the Recipient earning $70,000 is

---

anced Reform in Row 7, they are indifferent between ($500, −$500) and ($912.5, −$1087.5), where the latter is the mid-point of ($875, −$1125) and ($950, −$1050). Plugging these reforms into Equation (4), we get $\tilde{g} = \frac{\tilde{g}_h}{\tilde{g}_l} = \frac{rt-500}{t-500} = \frac{912.5-500}{1087.5-500} = 0.702$.

[12]The staircase method has several advantages. First, it is easy to explain to participants. Falk et al. (2018) use the staircase method in nationally representative samples across the world, highlighting its simplicity. Second, it allows us to get more accurate weights with fewer decisions than a typical multiple-price list because it adaptively selects a few questions from a list.

common across the six decision screens, a Social Architect's choices across the six decision screens are used to identify the relative welfare weights assigned to the seven Recipients.[13] We re-normalize the weights so that their sum is 1. As a result, the weight assigned to an individual can be interpreted as the share of the total weight.
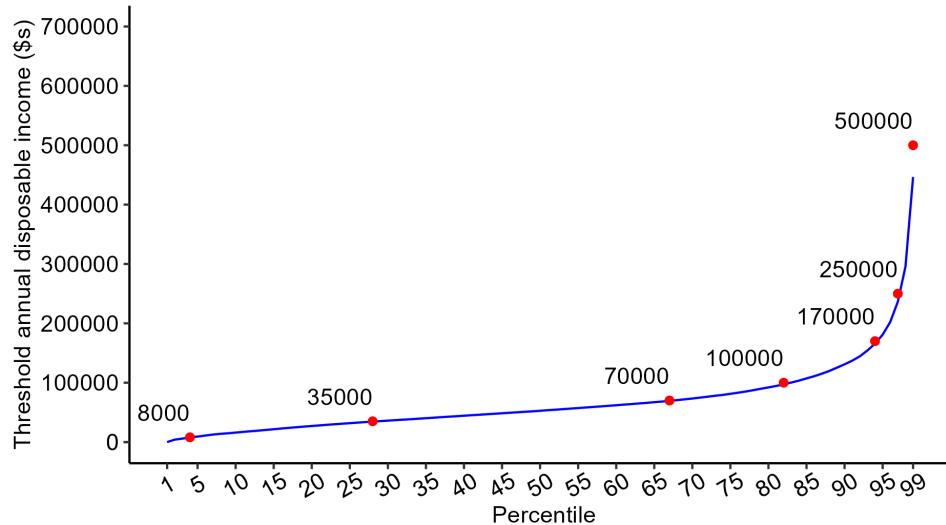

Figure 2: Disposable Incomes of the Seven Recipients

Notes: The figure plots the incomes of the seven Recipients (dots) against the disposable income distribution (line) in the U.S. in 2019. The horizontal axis indicates the percentiles and the vertical axis indicates the threshold annual disposable incomes corresponding to the percentiles. Data on income is obtained from the World Inequality Database (WID). Details on the construction of the figure can be found in Appendix Section E.1.

Incentives

We informed the Social Architects that one of them will be randomly selected in the study. For the selected Social Architect, one randomly selected choice in one randomly selected decision screen will be implemented. Thus, at the end of the study, two Recipients would receive $(\$1500 + rt, \$1500 - t)$ or $(\$1500 + 500, \$1500 - 500)$, depending on the choice of the selected Social Architect.

We included a sentence at the top of each decision that reminded Social Architects of the incentives. Figure A1 in Appendix Section C.1 presents a screenshot of one of the decisions in the experiment.

---

[13]A simple example can illustrate the calculation. Consider three Recipients: R1 (lowest income), R2, and R3 (highest income). A Social Architect has $\tilde{g} = 0.25$ when facing R1 and R3 and $\tilde{g} = 0.5$ when facing R2 and R3. Set the welfare weight assigned to R3 as 1. The implied welfare weight on R1 is $\frac{1}{0.25} = 2$, and R2 is $\frac{1}{0.5} = 4$. Re-normalizing these weights to sum to 1 gives us the welfare weights assigned to R1, R2, and R3, as $\frac{2}{7}, \frac{4}{7}$, and $\frac{1}{7}$.

14

Information about Recipients

A Social Architect learns that they will face seven real Recipients who will be randomly selected from a survey panel and will not participate in the same survey as them. They learn that the Recipients are above the age of 18 and are U.S. citizens. They view the after-tax incomes of the seven Recipients. Finally, we inform them about the incentives described in the previous section. We refer to the Social Architects as "Participants" and the Recipients as "Persons." We refer to the annual disposable incomes of the Recipients as "After-tax annual income."

3.2 Treatments and Waves

We collected data over two waves of data collection. In Wave 1 of data collection, we recruited 1965 participants in the role of Social Architects. These participants were randomly assigned to one of four treatments designed to test the robustness of the elicited welfare weights with respect to changes in the experimental design. In Wave 2, we recruited 1992 participants in the role of Social Architects. We implemented four treatments in Wave 2 to further test whether the welfare weights are robust to changes in the experimental design.

In Wave 1 of data collection, Social Architects are randomly assigned to one of four treatments in a 2×2 design. The first dimension tests if Social Architects' welfare weights are sensitive to the framing of the reforms. While Treatments Loss involve taking money away from the higher-income Recipient and giving money to the lower-income Recipient, Treatments Gain involve giving money to both Recipients in the pair. In Treatments Loss, each Recipient is given an initial endowment of \$1,500, and a Social Architect decides between the reforms ($\$rt, -\$t$) and ($\$500, -\$500$). In Treatments Gain, there is no initial endowment, and a Social Architect decides between ($\$1500 + rt, \$1500 - t$) and ($\$2000, \$1000$). The experimental design described in the previous sections corresponds to Treatments Loss. The framing of the reform should not affect a Social Architect's assigned welfare weights since the welfare weights depend on Recipients' consumption, which is the same in the two treatments. However, if a Social Architect is influenced by loss aversion, then, in Treatments Loss, they will choose Unbalanced Reforms with smaller values of t, assigning less progressive weights.

The second dimension tests if Social Architects' welfare weights are sensitive to the income of the Recipient common across the decision screens. In Treatments 70K, the Recipient common across the decision screens has an income of \$70,000. In contrast, in Treatments 500K, the Recipient common across the decisions screens has an income of \$500,000. The experimental design described in the previous sections corresponds to

Treatments 70K. We test whether the welfare weights elicitation is sensitive to the choice of the Recipient common across the decision screens.

In Wave 2 of data collection, Social Architects are randomly assigned to one of four treatments. The first two treatments test whether Social Architects' welfare weights are sensitive to the existence of real stakes (albeit small stakes). In Treatment Real, Social Architects make real decisions regarding real Recipients, while in Treatment Hypothetical, Social Architects make hypothetical decisions regarding hypothetical Recipients. If the welfare weights from the two treatments are similar, then future research can use hypothetical decisions. These are cheaper and easier to implement. However, if the welfare weights from the two treatments differ, then we defer to Treatment Real since the presence of real stakes is likely to lead to more reliable welfare weights. Both treatments are similar to Treatments Loss $\times$ 70K in structure.[14]

In the above treatments, Social Architects assume the role of impartial spectators, with self-interest motives playing a minimal role. However, in reality, people are often impacted by reforms. To examine the role of self-interest motives, we included Treatments No Self-Interest and Self-Interest. In Treatment No Self-Interest, Social Architects act as impartial spectators. In Treatment Self-Interest, Social Architects can potentially be affected by their own choices, meaning that one of them could potentially receive a payment in the study.[15] Treatment Self-Interest differs from Treatment Real in two ways. First, in Treatment Self-Interest, instead of viewing the exact incomes of the seven Recipients, Social Architects view the income brackets of the seven Recipients. This feature allows us to assign every Social Architect to one of the seven distinct income brackets that span the income distribution. Second, in Treatment Self-Interest, each Social Architect replaces the Recipient whose income bracket contains their own income. For example, if a Social Architect earns $400,000, they would replace the first Recipient whose income bracket is "$375,000 and above." Figure A2 in Appendix Section C.1 presents a screenshot of a decision faced by a Social Architect earning $400,000 in Treatment Self-Interest. Treatment No Self-Interest is similar to Treatment Self-Interest in that Social Architects view the income brackets of the seven Recipients rather than their exact incomes. However, it differs from Treatment Self-Interest in that Social Architects in Treatment No Self-Interest cannot be affected by their own choices.

---

[14]The only difference between these treatments is that while we referred to the incomes of the Recipients as "after-tax" in all treatments in Wave 1, in Treatment Real and Hypothetical, we additionally mentioned that the after-tax incomes of the Recipients were "accrued after all taxes and transfers."

[15]We did not inform the Social Architects in this Treatment whether they could potentially receive a payment based on the decision of another selected Social Architect.

Table 3: Overview of Treatments

Wave 1		
Treatment	Framing of Reforms	Income of Common Recipient
Loss × 70K	Loss	70K
Gain × 70K	Gain	70K
Loss × 500K	Loss	500K
Gain × 500K	Gain	500K
Wave 2		
Treatment	Framing of Reforms	Income of Common Recipient
Real	Loss	70K
Hypothetical	Loss	70K
No Self-Interest	Loss	70K
Self-Interest	Loss	70K

3.3 Additional Questions

We present an overview of the additional questions that Social Architects are asked to answer. The wording of the questions can be found in Appendix Sections F and G.

Wave 1

We elicit Social Architects' support for government redistribution by asking them (i) their views on the taxes levied on those in the top-income tax category and (ii) whether the government should reduce income differences between the rich and the poor. We use these two questions to test whether Social Architects' welfare weights predict their support for government redistribution.

Wave 2

We elicit Social Architects' support for government redistribution using a different question. The question asks them to consider the current incomes of individuals in society after all taxes and transfers and state whether they would like to redistribute incomes further.

To explore the predictors of Social Architects' support for redistribution, we elicit their misperceptions about taxation and society and views about taxation and government.

To understand whether Social Architects' welfare weights reflect their underlying ideals, we present them with three sets of questions. First, we elicit Social Architects' confidence in their decisions. Second, we elicit Social Architects' beliefs about whether high-income individuals and low-income individuals need and deserve their current incomes. We use these questions to test whether Social Architects' welfare weights capture ideals

related to Recipients' needs or factors unrelated to Recipients' needs. Third, we elicit Social Architects' beliefs about the source of income. We test whether Social Architects' welfare weights capture their beliefs about the source of income.

3.4 Data Collection

In Wave 1 of data collection, we recruited participants in the role of Social Architects from the data collection provider Lucid.[16] The collected sample includes participants from the general population of the U.S. Participants first answer questions about their demographics and political affiliation. We define quotas for recruitment based on gender, age, education, individual income, and region. The quotas are designed to match the sample to the population of the U.S. We exclude participants arriving after the quotas are met. Next, participants answer a question that serves as an attention check. Participants who fail the attention check are dropped from the study. Participants who pass the attention check are randomly assigned to one of the four treatments. After being assigned to the treatments, participants view the instructions and are asked to answer two questions that test their comprehension of the instructions. Participants who answer either of the two questions incorrectly are dropped from the study. We implemented the survey using Qualtrics. The data collection for Wave 1 began on 8 December 2021 and lasted approximately two weeks. Our final sample includes 1965 participants.[17]

In Wave 2 of data collection, we recruited participants in the role of Social Architects from the data collection provider Prolific.[18] The recruitment procedure is similar to the procedure used in Wave 1, except that in Wave 2, we do not implement any quotas during the recruitment stage. We implemented the survey using oTree (Chen et al., 2016). The data collection for Wave 2 began on 14 December 2022 and lasted eight days. Our final sample includes 1992 participants.[19]

---

[16]Lucid is commonly used in the literature (e.g., Haaland and Roth, 2023; Haaland et al., 2023).

[17]We recruited 6,735 participants in Wave 1. After dropping participants with multiple survey responses (0.09%), who did not consent to participate in the study (2.5%), who did not fit into one of the demographic quotas or did not reside in the U.S. (28.1%), who dropped out before the attention check (6.4%), who failed the attention check (28.2%), who dropped out before the comprehension check (16.5%), who failed the comprehension check (21.3%), who dropped out after passing the comprehension check (5.8%), we are left with 1965 participants. The share of participants that passed the attention check but dropped out before the comprehension check is not different across the four treatments ($F = 0.5228$, $p = 0.66$). The share of participants that passed the comprehension checks is not different across the four treatments ($F = 1.763$, $p = 0.1522$).

[18]Prolific has been used in several recent studies (e.g., Bursztyn et al., 2023; Enke et al., 2023). In Wave 2, we used Prolific instead of Lucid because Prolific allows us to pay participants a bonus, a feature required in Treatment Self-Interest. This feature is not available in Lucid.

[19]We recruited 2,313 participants in Wave 2. After dropping participants with multiple survey responses (0.3%), who did not consent to participate in the study (0.04%), who dropped out before the attention check (1.7%), who failed the attention check (2%), who dropped out before the comprehension check (3.1%), who failed the comprehension check (5.9%), who dropped out after passing the comprehension check (1.6%),

3.5 Summary Statistics

Table 4 presents the average characteristics of our sample and the population of the U.S. in 2019. The sample characteristics in Wave 1 closely match the population characteristics because we implemented quotas while recruiting participants.[20] The sample characteristics in Wave 2 also broadly match the population characteristics.[21] In the following sections, we report analyses that are weighted using sampling weights that ensure that the sample averages match the population averages.

Table A1 and Table A2 in Appendix Section C.2 presents the average characteristics of the sample across the four treatments in Wave 1 and Wave 2, respectively. We find statistically significant differences for several characteristics across treatments, but the magnitude of the differences for most of these characteristics is small. When we explore the welfare weights across treatments, we account for these imbalances by weighting each treatment using sampling weights such that the sample averages in each treatment match the population averages.

3.6 Pre-registration

We pre-registered the design as well as the analyses. There are minor deviations from the pre-registration in the implementation of the experiment. There are a few deviations from pre-registration in the analyses. The deviations are discussed in Appendix Section B.

4 Welfare Weights of the General Population

4.1 Data Description

We explore the distribution of $\tilde{g}$, the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, across the six pairs of Recipients in the six decision screens. The parameter $\tilde{g}$ is calculated using Equation (4).

Figure 3 presents the results, each sub-figure relating to a pair of Recipients. Each of the sixteen bars in the figure corresponds to a unique value of $\tilde{g}$, which is derived from the unique row in Table 1 where the Social Architects switch from a Balanced Reform to an Unbalanced Reform. The last bar corresponds to those who never switch. The figure uses

---

we are left with 1992 participants. The share of participants that passed the attention check but dropped out before the comprehension check is not different across the four treatments ($F = 0.7024, p = 0.5506$). The share of participants that passed the comprehension check is not different across the four treatments ($F = 0.168, p = 0.918$).

[20]Because we relaxed the quotas towards the end of the study to expedite reaching our target sample size, our sample has a higher share of people with education up to high school compared to the population.

[21]Our sample has a lower share of individuals with incomes below $30,000, a lower share of individuals above the age of 64, a lower share of individuals who have studied up to high school, a higher share of individuals with a bachelor's degree, and a lower share of Republicans.

Table 4: Summary Statistics

	Population	Wave 1	Wave 2
Income: < 30,000	0.51	0.53	0.38
Income: 30-59,999	0.26	0.26	0.29
Income: 60-99,999	0.14	0.13	0.22
Income: 100-149,999	0.06	0.05	0.09
Income: > 149,999	0.04	0.03	0.04
Age: 18-34	0.30	0.29	0.37
Age: 35-44	0.16	0.17	0.22
Age: 45-54	0.16	0.17	0.15
Age: 55-64	0.17	0.17	0.16
Age: > 64	0.21	0.19	0.10
Edu: Up to Highschool	0.39	0.46	0.14
Edu: Some college	0.22	0.20	0.20
Edu: Bachelor or Associate	0.28	0.24	0.49
Edu: Masters or above	0.11	0.10	0.16
Region: West	0.24	0.21	0.18
Region: North-east	0.17	0.18	0.20
Region: South	0.38	0.40	0.43
Region: Mid-west	0.21	0.21	0.20
Male	0.49	0.46	0.50
Republican	0.28	0.32	0.19

Notes: Population average demographics are computed using the 2019 American Community Survey (ACS) 1-year estimates. The ACS sample only includes individuals above the age of 18. The population share of Republicans is obtained as the average share of people identifying as Republican over multiple surveys fielded in 2019 by the Gallup poll (https://news.gallup.com/poll/15370/party-affiliation.aspx). The sample means from Wave 1 are computed using the 1965 Social Architects recruited from Lucid. The sample means from Wave 2 are computed using the 1992 Social Architects recruited from Prolific.

the data of Social Architects in all treatments and waves where the Recipient common across the pairs of Recipients has an income of \$70,000.

We observe considerable heterogeneity in $\tilde{g}$ across the pairs of Recipients, comprising both progressive welfare weights ($\tilde{g} < 1$) and regressive welfare weights ($\tilde{g} > 1$).[22] This heterogeneity also extends within these two groups. The distribution of $\tilde{g}$ across the pairs of Recipients is skewed towards progressivity (towards $\tilde{g} = 0$), suggesting that the welfare weights assigned to the seven Recipients are progressive on average.

---

[22]A value $\tilde{g} < 1$ corresponds to a lower weight assigned to the higher-income Recipient, which implies that the welfare weights assigned to the two Recipients are progressive. A value $\tilde{g} > 1$ corresponds to a higher weight assigned to the higher-income Recipient, which implies regressive weights.

We observe bunching across the six pairs of Recipients at three values: $\tilde{g} = 0.03$, representing the most progressive weights; $\tilde{g} = 0.9$, representing the least progressive among the progressive weights; and $\tilde{g} = 39$, representing the most regressive weights.[23] The share bunching at $\tilde{g} = 0.9$ is much smaller than the share bunching at the other two values of $\tilde{g}$.

We observe a similar shift in the distribution of $\tilde{g}$ towards progressivity and similar bunching in the treatments where the Recipient common across the pairs of Recipients has an income of $500,000. The results are presented in Figure A3 in Appendix Section C.1.
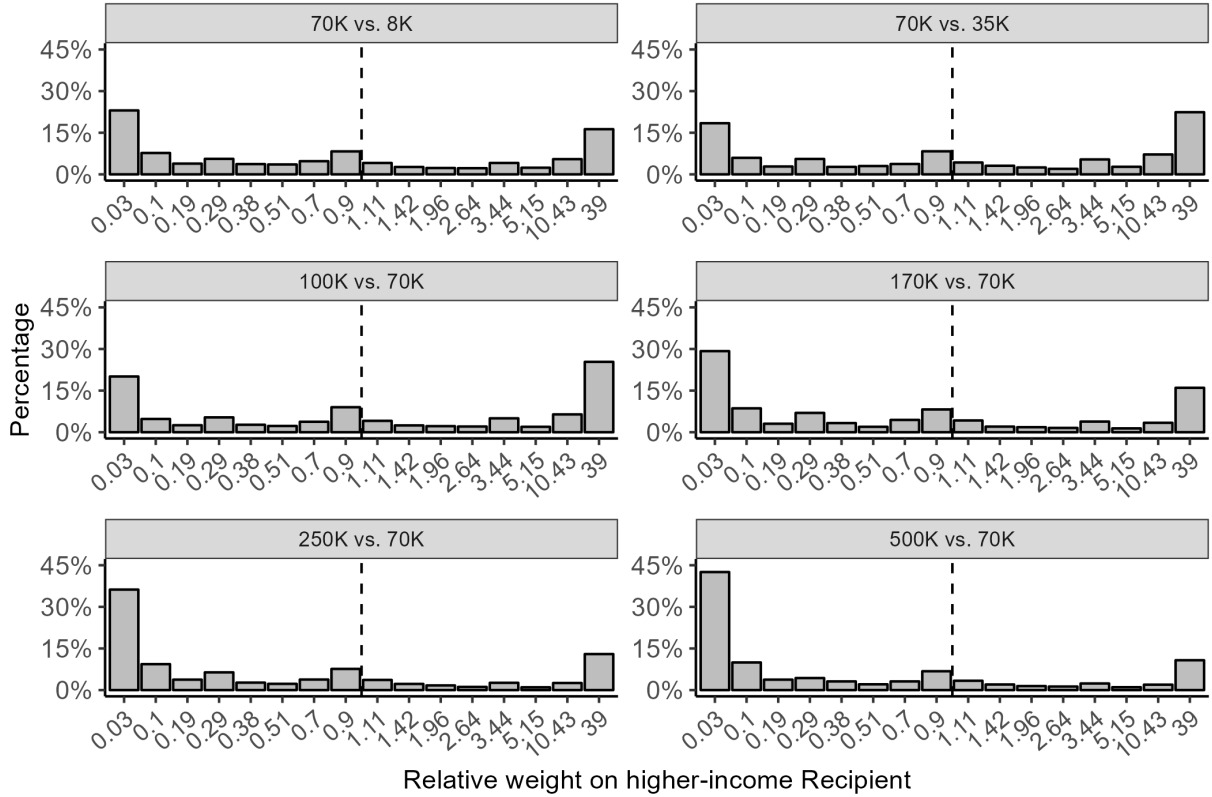

Figure 3: Distribution of $\tilde{g}$ Across Recipient Pairs

Notes: The figure presents the distribution of $\tilde{g}$, the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, across the six pairs of Recipients in the six decision screens. The parameter $\tilde{g}$ is calculated using Equation (4). The sample includes Social Architects in all treatments in both waves where the Recipient common across the pairs of Recipients has an income of $70,000.

---

[23]The value $\tilde{g} = 0.03$ and $\tilde{g} = 39$ are observed when a Social Architect selects the Unbalanced Reform and Balanced Reform, respectively, in all four decisions. The value $\tilde{g} = 0.09$ is observed when a Social Architect selects the Unbalanced Reform in the first decision and the Balanced Reform in the subsequent three decisions.

4.2 Progressivity of Welfare Weights

For each Social Architect, we estimate their assigned $\tilde{g}$ across the six decision screens. Since there is a common Recipient across the screens, the set of $\tilde{g}$ values for each Social Architect is used to identify $g(R_1) \ldots g(R_7)$, the relative welfare weights assigned to the seven Recipients. In this section, we explore the average progressivity of the welfare weights across all Social Architects. We discuss issues in aggregation in Section 6.

To estimate the progressivity of the welfare weights, we use a parametric function that is often used in the literature (e.g., Allcott et al., 2019; Saez, 2002). In particular, we identify the parameter ν that makes the function y^ν a best fit of the welfare weights, where y is the disposable incomes of the Recipients, and ν is a parameter that governs the progressivity of the weights. The parameter ν can be interpreted as the elasticity of the welfare weights with respect to Recipients' incomes. A value $\nu < 0$ indicates progressive weights, while a value $\nu > 0$ indicates regressive weights.

Using the simple parametric function y^ν allows the general population weights estimated in our paper to be portable to other contexts. Optimal policy formulas based on Recipients' incomes can use the function y^ν, where ν is estimated based on the general population weights. Optimal policy formulas based on Recipients' utilities can characterize welfare weights using constant relative risk aversion (CRRA) utilities. In particular, there is a mapping from CRRA utilities with a coefficient of relative risk aversion denoted by γ and the function y^ν given by $\gamma = -\nu$.[24]

To estimate the parameter ν, we estimate the following linear regression.

$$log(g(R_j)_i) = \beta_0 + \nu log(recipient\ income_j) + \epsilon_{ij} \tag{5}$$

where $log(g(R_j)_i)$ is the natural logarithm of the welfare weight assigned by Social Architect i to Recipient j and $recipient\ income_j$ is the disposable income of Recipient j. We estimate the regression using Social Architects from all the treatments in both waves. We weight this regression using sampling weights constructed such that the average sample characteristics match the population characteristics.[25]

We find that the estimated value of ν is -0.34, implying progressive welfare weights on average. Our elasticity estimate implies that if a Social Architects assign a weight of

---

[24]CRRA utilities are given by $u(c,l) = u(c) + v(l) = \frac{c^{1-\gamma}-1}{1-\gamma} - v(l)$, where c represents consumption, v represents the disutility of work (which depends on labor effort l), and γ is the coefficient of relative risk aversion. Taking the first-order condition of the utilitarian social welfare function $W = \sum u(c,l)$ results in $\Delta W = \sum u'(c)$. We have $\Delta W = c^{-\gamma}\Delta c$, where $u'(c) = c^{-\gamma}$ characterizes the welfare weights.

[25]The average population characteristics are taken from Column (1) in Table 4.

1 to a Recipient, they would assign a weight of 66 cents to a Recipient earning twice as much. This means that for every taken from a Recipient, Social Architects are willing to give as little as 66 cents to a Recipient earning half as much. This leads to the following result.

Result 1. *The general population weights are progressive on average.*

Since the welfare weights were elicited given the current tax and transfer system, our finding that $v < 0$ suggests that, on average, the general population wants additional redistribution beyond that achieved by the current tax and transfer system.[26]

We compare our elasticity estimate to the benchmarks proposed in the literature: $v = |0.25|$ corresponds to "weak" redistributive tastes, $v = |1|$ corresponds to "fairly strong" redistributive tastes, and $v = |4|$ corresponds to "extremely strong" redistributive tastes (Allcott et al., 2019; Saez, 2002). Based on our experimental design, the range of the possible values of v is $[-2.25, 2.25]$. Since only 15% of the Social Architects make choices that imply the most regressive or most progressive weights, the limited range of v does not severely affect our inferences. Comparing our estimate of -0.34 to the benchmarks suggested in the literature suggests that the general population weights are characterized by weak to strong redistributive tastes.

4.3   Individual Heterogeneity in Welfare Weights

In this section, we explore the individual-level heterogeneity in the progressivity of the welfare weights. We estimate the elasticity of the weights with respect to Recipients' incomes (v) for each Social Architect using the following regression.[27]

$$log(g(R_j)) = \beta_0 + vlog(recipient\ income_j) + \epsilon_j \tag{6}$$

where $g(R_j)$ is the welfare weight assigned by a Social Architect to Recipient j and *recipient income$_j$* is the disposable income of Recipient j.

We find 65% of the Social Architects have progressive welfare weights ($v < 0$), while the remaining 35% have regressive welfare weights ($v > 0$). A share of the Social Archi-

---

[26]A value $v = 0$ would imply an indifference between giving a dollar to the rich and to the poor, indicating that the general population is satisfied with the redistribution achieved by the current system.

[27]We estimate the elasticity of Social Architects' welfare weights instead of exploring the share of participants with weakly monotonic weights, i.e., weights that are weakly increasing or decreasing with the incomes of the Recipients, because only 23% of the participants in our study have weakly monotonic weights. Non-monotonic weights do not necessarily imply measurement error in the weights. Instead, they can be consistent with various underlying ideals. For example, Social Architects may have progressive weights with a downward spike in the weight assigned to very low-income individuals, who they believe are "lazy" (Drenik and Perez-Truglia, 2018).

tects with regressive weights may be libertarians, advocating for no additional redistribution.[28] Our results align with findings from Almås et al. (2020), who find, in a different setting, that 30% of the participants sampled from a nationally representative population of the U.S. are libertarians.

Only about 15% of the Social Architects make choices that imply the most regressive (5%) or most progressive (10%) welfare weights, which suggests that there is limited polarization in Social Architects' welfare weights. There is considerable heterogeneity in the estimated values of ν, which can be seen in the cumulative distribution function of the estimated values of ν presented in Figure A4 in Appendix Section C.1.[29]

We compare the estimated values of ν with the benchmarks in the literature. We group those with progressive weights and regressive weights. About 46% of the Social Architects exhibit weak redistributive tastes, 41% have strong redistributive tastes, and the remaining 13% exhibit extremely strong redistributive tastes. Table A3 in Appendix Section C.2 presents the results.

4.4 Welfare Weights and Ideals

Is the heterogeneity in welfare weights due to heterogeneous underlying ideals? We present three results to support the hypothesis that Social Architects' welfare weights are driven by their underlying ideals.

First, we assess Social Architects' confidence in their decisions. In Wave 2, we elicited Social Architects' confidence in their decisions. Their responses to this question could range from "1: Not confident at all" to "5: Completely confident." We hypothesized that Social Architects who made mistakes, either because of a trembling-hand mistake or a poor understanding of the task, would report lower confidence levels. We find that 83% of the Social Architects reported a confidence level of 4 or 5. This result suggests that the heterogeneity in Social Architects' welfare weights is likely due to their underlying ideals rather than mistakes.[30]

Second, we test whether Social Architects' welfare weights capture "non-welfarist" ideals. In the traditional "welfarist" approach, welfare weights depend on characteristics that enter Recipients' utilities (e.g., Recipients' disability status). On the other hand, in the non-welfarist approach, welfare weights do not enter Recipients' utilities (e.g., Recip-

---

[28]Some Social Architects with regressive weights might favor reducing the redistribution due to the current tax system. However, they may not align with the libertarian principle of no redistribution.

[29]We observe similar heterogeneity in the slope of the welfare weights assigned by Social Architects with respect to the index of the Recipients (1 through 7). The results are presented in Figure A5 in Appendix C.1.

[30]We find no evidence that Social Architects with high confidence (above median confidence) assign different welfare weights relative to those with low confidence. This result suggests that the welfare weights are unlikely to be severely biased by the presence of Social Architects with low confidence levels.

ients' parental income). Non-welfarist ideals include a wide range of normative ideals, such as equality of opportunity or distribution based on the source of income (Saez and Stantcheva, 2016). Evidence of welfare weights capturing non-welfarist ideals would suggest that the heterogeneity in welfare weights likely reflects heterogeneous non-welfarist ideals. We implemented two questions in Wave 2 to identify if welfare weights capture welfarist or non-welfarist ideals. Our results suggest that the Social Architects' welfare weights capture both welfarist and non-welfarist ideals. The results are presented in Appendix Section D.1.

Third, we test whether Social Architects' welfare weights capture an important non-welfarist ideal documented in the literature: redistribution based on the source of income. Several studies have documented that people are willing to accept inequalities when incomes are earned due to effort but not if they are earned due to luck (e.g., Almås et al., 2020). In Wave 2, we asked Social Architects their beliefs about whether high-income individuals are rich due to luck or effort. Social Architects believing that high-income individuals earned their income through effort assign less progressive welfare weights than those attributing high incomes to luck. Our finding that welfare weights capture an important normative ideal documented in the literature provides suggestive evidence that the heterogeneity in welfare weights likely reflects underlying ideals.

The above three results suggest that the heterogeneity in Social Architects' welfare weights is likely due to the heterogeneous underlying ideals. This supports our general sufficient-statistics approach to eliciting welfare weights that can allow Social Architects to be guided by various welfarist and non-welfarist ideals.

4.5 Welfare Weights and Background Characteristics

In this section, we explore the role of Social Architects' background characteristics in explaining the observed heterogeneity in welfare weights, focusing on two important characteristics highlighted in the literature: income and political affiliation. To explore the role of background characteristics, we estimate the following regression.

$$log(g(R_j)_i) = \beta_0 + \nu_0 log(recipient\ income_j) + \beta_1 x_i^1 + \dots \beta_n x_i^n + \tag{7}$$
$$\nu_1 x_i^1 * log(recipient\ income_j) + \dots + \nu_n x_i^n * log(recipient\ income_j) + \epsilon_{ij}$$

where $log(g(R_j)_i)$ is the natural logarithm of the welfare weight assigned by Social Architect i to Recipient j and $recipient\ income_j$ is the disposable income of Recipient j. The variables $x^1, \dots, x^n$ (shortened to X) include a set of treatment dummies from Wave 1 and Wave 2 and Social Architects' background characteristics including *Republican* (=1 if

Republican), *High Income* (= 1 if above median income), *Male* (=1 if male), *High Education* (=1 if above median education), and *High Age* (=1 if above median age).

Figure 4 presents the coefficient estimates. Except for the main effect of the log of the incomes of the Recipients, we do not present the main effects of the other variables. We also do not present the coefficient estimates related to the treatment dummies. The coefficient estimate of $log(recipient\ income_j)$ indicates the estimated elasticity of the weights in the base category or base characteristic. The coefficient estimates of $log(recipient\ income) \times$ *Characteristic* can be interpreted as the change in the estimated elasticity in the given characteristic relative to the elasticity in the base characteristic.

The coefficient estimate in the second row is 0.24, which implies that Republicans assign less progressive weights relative to Democrats and Independents. Our results suggest that a part of the partisan gap in support for redistribution documented in the literature (Fisman et al., 2017; Singhal, 2008) is likely driven by the partisan gap in welfare weights. We find that Social Architects' welfare weights explain 8% of the partisan gap in preferences for redistribution at the margin. The results are presented in Appendix Section D.5. While we see a partisan gap in welfare weights, we find that, on average, Republicans assign progressive welfare weights, implying that they also want additional redistribution at the margin.

Social Architects with higher incomes, i.e., above median incomes ($32,700), have less progressive weights (ν is 0.16 higher). Our results suggest that a part of the gap in support for redistribution between low and high-income individuals documented in the literature (Cohn et al., 2019; Singhal, 2008) is likely due to the gap in welfare weights. We also investigate whether Social Architects assign a higher weight to Recipients with similar incomes to their own compared to other Recipients. Our results suggest that Social Architects assign a higher weight to Recipients with incomes similar to their own. This effect is larger for Social Architects with lower incomes. We present the results in Appendix Section D.2.

We conduct various robustness checks to explore the role of demographics that are presented in the appendix. We find similar results in these robustness checks.[31] We now have the following result.

Result 2. *Social Architects with higher incomes and those identifying as Republicans assign less progressive welfare weights.*

---

[31]We estimate linear regressions in which the dependent variable is the individual-level elasticity of Social Architects' weights with respect to Recipients' incomes (obtained from Equation (6)). Table A5 in Appendix Section C.2 presents the regression estimates. The results from these regressions are similar to those found in Figure 4.
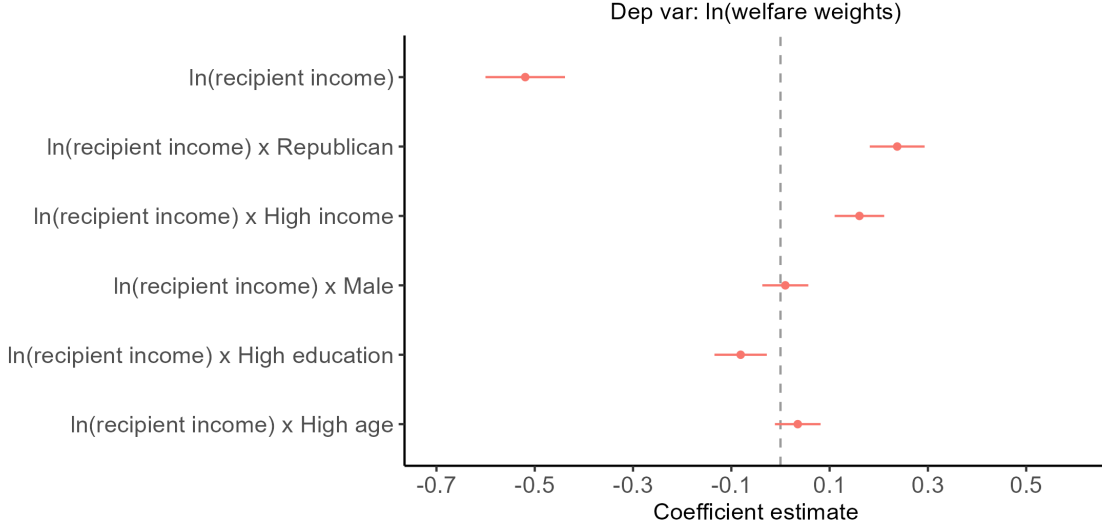
Figure 4: Social Architects' Welfare Weights and Characteristics

Notes: The figure presents coefficient estimates. The dependent variable is the log of the welfare weights assigned by Social Architects. The explanatory variables are the log of the incomes of the Recipients, a set of treatment dummies, and Social Architects' characteristics including *Republican* (=1 if Republican), *High Income* (= 1 if above median income), *Male* (=1 if male), *High Education* (=1 if above median education), and *High Age* (=1 if above median age), and the interaction terms of the log of the incomes of the Recipients with all the other variables. Except for the main effect of the log of the incomes of the Recipients, we do not present the main effects of the other variables. We also do not present the coefficient estimates related to the treatment dummies. The regressions are weighted using sampling weights. Error bars are computed using HC3 standard errors.

4.6 Treatment Effects

We implemented four treatments in Wave 1 and four treatments in Wave 2 to test whether the welfare weights are robust to changes in the experimental design. In this section, we compare the progressivity of Social Architects' welfare weights across treatments. We estimate the regression specified in Equation (5) separately for each treatment. We present unweighted and weighted regressions but focus on the weighted regressions. The weighted regressions use sampling weights constructed such that the average characteristics in each treatment match the average population characteristics. Table 5 presents the results.

Role of Framing of the Reforms: To test whether Social Architects' welfare weights are sensitive to the framing of the reforms, we compare Treatments Loss × 70K to Gain × 70K. In Treatment Loss × 70K, participants face reforms that take money away from the higher-income Recipient and give money to the lower-income Recipient. In contrast, in Treatment Gain × 70K, the reforms are framed as a gain to both Recipients. Looking at the second and third rows in Column (3) of the table, we find that Social Architects have

27

more progressive welfare weights in Treatment Gain $\times$ 70K relative to Treatment Loss $\times$ 70K (v is 0.14 lower), and this effect is statistically significant.[32] In Treatments Loss, Social Architects may be reluctant to take money away from the higher-income Recipients due to loss-aversion (**?**). A higher reluctance to take money away from the higher-income Recipient would lead to less progressive welfare weights.

We also test the role of gain framing by comparing Treatment Loss $\times$ 500K to Treatment Gain $\times$ 500K. We find that Social Architects have more progressive welfare weights in Treatment Gain $\times$ 500K relative to Treatment Loss $\times$ 500K (v is 0.04 lower). This effect is smaller than the effect we found above. Furthermore, this effect is not statistically significant.

Role of the Common Recipient: To test whether Social Architects' welfare weights are sensitive to the income of the Recipient common across the decision screens, we compare Treatments Loss $\times$ 70K and Loss $\times$ 500K. In the former, the Recipient common across the decision screens has an income of $70,000, while in the latter, the Recipient common across the decision screens has an income of $500,000. We find that Social Architects have less progressive welfare weights in Treatment Loss $\times$ 500K relative to Loss $\times$ 70K (v is 0.27 higher). A part of the treatment difference is a mechanical effect. In Treatment Loss $\times$ 500K, always choosing the Unbalanced Reform would imply equal weights for Recipients one through six because the first six Recipients are compared to the seventh Recipient and because there is a bound on the welfare weights. This results in a mechanical flattening of the welfare weights.[33] If Social Architects choose the Unbalanced Reform in every decision in every decision screen, the elasticity of the weights in Treatment Loss $\times$ 70K is -2.25, while it is -0.567 in Treatment Loss $\times$ High, highlighting the flattening of the welfare weights in the latter treatment. However, a part of the treatment difference may also be due to a behavioral effect that may result from the different reference points in the two treatments.

We find a similar effect when we compare Social Architects in Treatment Gain $\times$ 70K to those in Treatment Gain $\times$ 500K (v is 0.36 higher).

Comparing Treatments Across Waves: Treatment Loss $\times$ 70K was implemented in Wave 1 of data collection, while Treatment Real was implemented in Wave 2 of data collection. These two treatments are similar in structure. We find that Social Architects are more

---

[32]To check whether the progressivity of the welfare weights is statistically significantly different between treatments, we present the results from a regression that includes a set of treatment dummies. The results can be found in Figure A6 in Appendix Section C.1.

[33]Although there is also a mechanical flattening of the weights in Treatment Loss $\times$ 70K, the extent of the flattening is lower because the welfare weights assigned to the Recipients earning less than $70,000 can be different from the welfare weights assigned to those earning more than $70,000.

Table 5: Elasticity of the Weights by Treatments

	(1)	(2)	(3)	(4)
Case	mean	se	mean	se
All	-0.35	0.01	-0.34	0.01
Loss x 70K	-0.36	0.03	-0.37	0.03
Gain x 70K	-0.49	0.03	-0.51	0.04
Loss x 500K	-0.09	0.03	-0.1	0.03
Gain x 500K	-0.13	0.02	-0.14	0.02
Real	-0.58	0.03	-0.53	0.05
Hypothetical	-0.67	0.03	-0.76	0.04
No Self-Interest	-0.14	0.03	-0.07	0.05
Self-Interest	-0.38	0.03	-0.46	0.04
Weighted?	No	No	Yes	Yes

Notes: The table presents the coefficient estimates (ν) and the standard errors of Social Architects' elasticity of the welfare weights with respect to Recipients' incomes. Each row presents regression estimates obtained by estimating Equation (5) using the subset of Social Architects indicated in the column "Case." In the first row, we pool Social Architects across all eight treatments. Columns (1) and (2) report the estimates from an unweighted regression, while Columns (3) and (4) report the estimates from a weighted regression using sampling weights. In the first row, the sampling weights are constructed such that the characteristics in the full sample match the population characteristics. In other rows, the sampling weights are constructed such that the characteristics in each treatment match the population characteristics.

progressive (ν is 0.15 lower) in Treatment Real relative to Treatment Loss $\times$ 70K. The difference is not due to differences in observables since we weight each treatment to match the population characteristics. However, we cannot rule out that the difference is due to unobservables: participants in the two waves may differ by unobservables. The patterns in the welfare weights across the two treatments are presented in Table A4 in Appendix Section C.2.

Role of Stakes: We examine whether the presence of real stakes affects Social Architects' welfare weights. In Treatment Real, Social Architects make real decisions regarding real Recipients, while in Treatment Hypothetical, Social Architects make hypothetical decisions regarding hypothetical Recipients. We find that Social Architects' welfare weights are more progressive in Treatment Hypothetical relative to Treatment Real (ν is 0.23 higher). We do not find evidence that participants in Treatment Hypothetical have more progressive welfare weights because they are less attentive in the survey. The results are presented in Appendix Section D.3.

Role of Self-Interest Motives: We explore the role of self-interest motives. In Treatment Self-Interest, Social Architects can potentially be affected by their own choices, while in Treatment No Self-Interest, Social Architects are in the role of impartial spectators. We find that participants in Treatment Self-Interest have more progressive weights than participants in Treatment No Self-Interest, indicating that Social Architects' welfare weights are guided by self-interest motives. We explore the role of self-interest separately for each income group in Appendix Section D.3.

Discussion: We find that framing the reforms as gains leads Social Architects to assign more progressive welfare weights, likely due to loss aversion. Social Architects assign less progressive weights when the Recipient common across the decision screens has a high income. This effect is partly a mechanical effect and partly a behavioral effect. Social Architects assign more progressive welfare weights in Wave 2 relative to Wave 1, likely because the pool of participants in Wave 2 is different on unobservables. Finally, we find that self-interest motives affect Social Architects' welfare weights. We compare Treatments Real and No-Self-Interest in Appendix Section D.3 and find that Social Architects are more progressive in Treatments Real.

The above results suggest that general population weights are characterized by a range of estimates. The range of general population weights would lead to a range of optimal policies. The range of policies that can be obtained using the general population weights is much narrower than the range obtained using the full range of welfare weights estimates in the experiment $\nu \in [-2.25, +2.25]$ and the range of estimates that are sometimes used in the literature $\nu \in [-4, +4]$ (e.g., Saez, 2002). We plot the elasticity estimates in each treatment with the x-axis ranging from -2.25 to 2.25 in Figure A7 in Appendix Section C.1.

4.7 Welfare Weights and Support for Redistribution

4.7.1 Do Welfare Weights Predict Support for Redistribution?

Social Architects' support for real-world government redistribution could depend on their welfare weights and other factors, such as beliefs about the behavioral responses to taxation. In this section, we explore the empirical link between Social Architects' welfare weights and their support for redistribution. This exercise highlights the value of using the welfare weights of the general population to identify socially acceptable policies.

In Wave 2, we elicited Social Architects' support for redistribution using a survey measure. The question asks Social Architects to consider the current incomes of individuals after all taxes and transfers and asks them if they want additional redistribution. Their responses can range from -2 to +2, where positive (negative) values indicate that income

should be further redistributed by taking from higher-income (lower/middle-income) individuals and giving to lower/middle-income (higher-income) individuals. This question has two key features: First, the question explicitly asks Social Architects to consider the current incomes of individuals in society after all taxes and transfers, thereby fixing Social Architects' beliefs about the status quo beyond which redistribution should occur. Second, the question allows regressive redistribution from the poor to the rich, which is useful to capture ideals such as libertarianism.

We estimate the correlation between Social Architects' elasticity of the weights and their support for redistribution. The elasticity of Social Architects' weights is computed using Equation (6). We find that the estimated correlation is -0.36, which is significant at the 1% level. This estimate implies that Social Architects with more progressive welfare weights have stronger support for progressive redistribution.

We benchmark the predictive power of Social Architects' elasticity of the weights against their stated political affiliation. Political affiliation has been identified as a consistent predictor of people's support for redistribution (e.g., Stantcheva, 2020). We regress Social Architects' support for redistribution on their elasticity of the weights or their political affiliation. To assess the predictive power of a specification, we compute the root mean squared error (RMSE) of the out-of-sample predictions obtained from a specification.[34] The lower the RMSE of a specification, the higher the predictive power of that specification.

Table 6 presents the results. Looking at the first two rows, we find that the elasticity of Social Architects' welfare weights is just as good a predictor of their support for redistribution as their stated political affiliation. We find similar results in Rows 3 and 4, which include treatment dummies and background characteristics as controls in the regressions.

We also find similar results when we conduct a benchmarking exercise based on two validated measures of support for redistribution that are used in the literature. The results are presented in Appendix Section D.4. This leads to the following result.

Result 3. *Social Architects' support for redistribution can be predicted with similar accuracy using either their stated political affiliation or their assigned welfare weights.*

Our findings above suggest that Social Architects' welfare weights predict their support for redistribution. This means that calibrating optimal policy formulas with the welfare weights of the general population would likely lead to socially acceptable policies.

---

[34]We divide the data into four sub-samples ($S(k), k \in 1, 2, 3, 4$) with $k = 4$. For each sub-sample, we train the specification of interest using the other three sub-samples ($S(-k)$). Next, we predict values for the sub-sample we left out and calculate the squared error, which is the difference between the actual and predicted values squared. To obtain the RMSE, we compute the square root of the average of the squared errors across all four sub-samples.

Table 6: Welfare Weights and Support for Redistribution

Row	Explanatory variable	Controls?	RMSE
1	Republican	No	0.85
2	Elasticity of the weights	No	0.83
3	Republican	Yes	0.84
4	Elasticity of the weights	Yes	0.83
5	Republican + Elasticity of the weights	No	0.80
6	Republican + Elasticity of the weights	Yes	0.80

Notes: Each row of the table presents the root-mean-squared error (RMSE) of the predictions generated from a linear regression using a k-fold cross-validation procedure with $k = 4$. The dependent variable (*Redistribution*) takes values from -2 to +2, where positive (negative) values indicate that income should be further redistributed by taking from higher-income (lower/middle-income) individuals and giving to lower/middle-income (higher-income) individuals. *Elasticity of the weights* is the elasticity of Social Architects' weights with respect to Recipients' incomes. *Republican* is a dummy variable taking a value of 1 for Republicans and a value of 0 for Democrats or Independents. The controls in the regression include a set of treatment dummies, *High Income* (= 1 if above median income), *Male* (=1 if male), *High Education* (=1 if above median education), and *High Age* (=1 if above median age). The regressions use data from Wave 2.

Looking at Rows 5 and 6 in Table 6, we find that in a specification that includes Social Architects' political affiliation as an explanatory variable, adding the elasticity of Social Architects' welfare weights in the regression improves the predictive power of the specification by a small extent. This result implies that the partisan gap in support for redistribution is partly driven by factors other than welfare weights. We explore the factors that explain the partisan gap in support for redistribution in Appendix Section D.5.[35]

The above exercise indicates that the progressivity of the welfare weights predicts the progressivity of support for government redistribution. How does the share of Social Architects who want additional redistribution based on their welfare weights compare to the share who support additional government redistribution in the survey measure? In Section 4.3, we found that 65% of the Social Architects have progressive welfare weights. Since the welfare weights were elicited given the current tax and transfer system, Social Architects with progressive welfare weights will, all else equal, desire additional redistribution. We find that 71% of the Social Architects support additional government redistribution in the survey measure. We present the frequencies of responses to this question in Figure A11 in Appendix Section C.1. The share of Social Architects who want addi-

---

[35]We explore the factors that explain the partisan gap in support for redistribution using the decomposition proposed by Gelbach (2016). We find that Social Architects' welfare weights explain 8% of the partisan gap in support for redistribution. The partisan gap is largely driven by concerns about externalities from inequality and concerns about the efficiency costs of taxation.

tional redistribution based on their welfare weights is similar to the share who support additional government redistribution in the survey measure.

4.7.2 Which Factors Predict Support for Redistribution?

We explore whether Social Architects' support for redistribution is driven by their welfare weights or other factors. We regress Social Architects' support for redistribution (explained in the previous section) on their standardized elasticity of welfare weights and other variables, capturing other factors, including Social Architects' misperceptions and views about taxation and government. Details about the other variables can be found in Appendix Section A.

Model s2 in Figure 5a presents the coefficient estimates of the regression. We find that Social Architects' elasticity of the weights predicts their support for redistribution. However, Social Architects' misperceptions also predict their support for redistribution. Social Architects who overestimate the level of taxes have less progressive support for redistribution. Intuitively, Social Architects who think that society already bears a high tax burden are less in favor of increasing the tax burden. Social Architects who overestimate upward mobility, i.e., who think income is more upward mobile than it is, have less progressive support for redistribution. This result is consistent with the findings of Alesina et al. (2018). Social Architects who overestimate the share of individuals earning less than $35,000 have more progressive support for redistribution.

Social Architects' views about taxes and government also predict their support for redistribution. Social Architects who think that higher taxes on high-income individuals hurt the economy have less progressive support for redistribution. Social Architects who believe in trickle-down economics have more progressive support for redistribution. However, the direction of this effect is counterintuitive. Finally, Social Architects who believe that inequality is a serious issue because it can have externalities have more progressive support for redistribution.

Overall, we find that while Social Architects' welfare weights predict their support for redistribution, their misperceptions and views about taxes and government also predict their support for redistribution. We have the following result.

Result 4. *Social Architects' welfare weights, misperceptions, and views about taxation and government jointly predict their support for redistribution.*

4.7.3 Which Factors do Welfare Weights Capture?

Our experimental measure of welfare weights is designed to capture welfare preferences. However, empirically, welfare weights may also capture factors orthogonal to welfare preferences. In this section, we explore which factors are captured by welfare

33
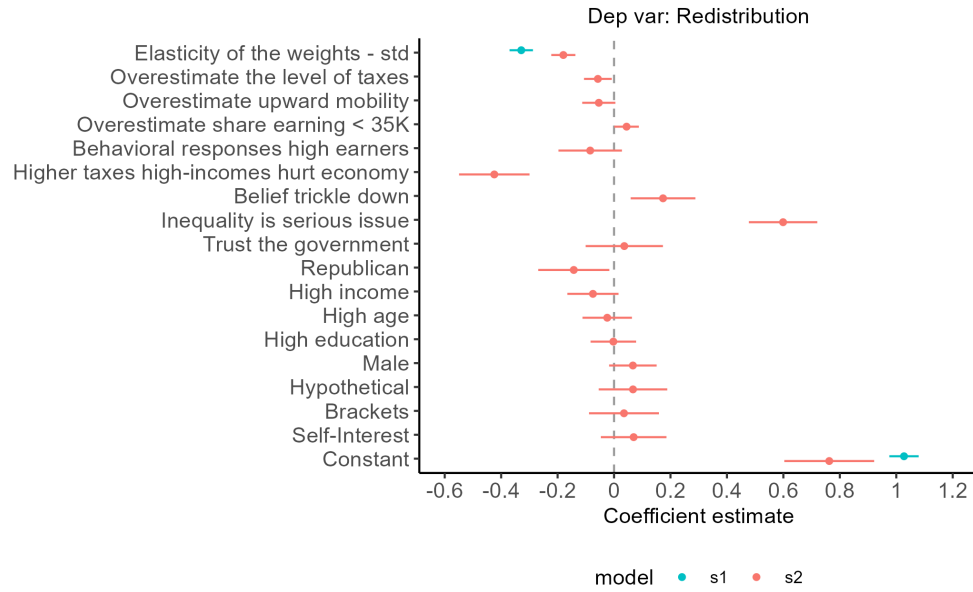
(a) Coefficient Estimates
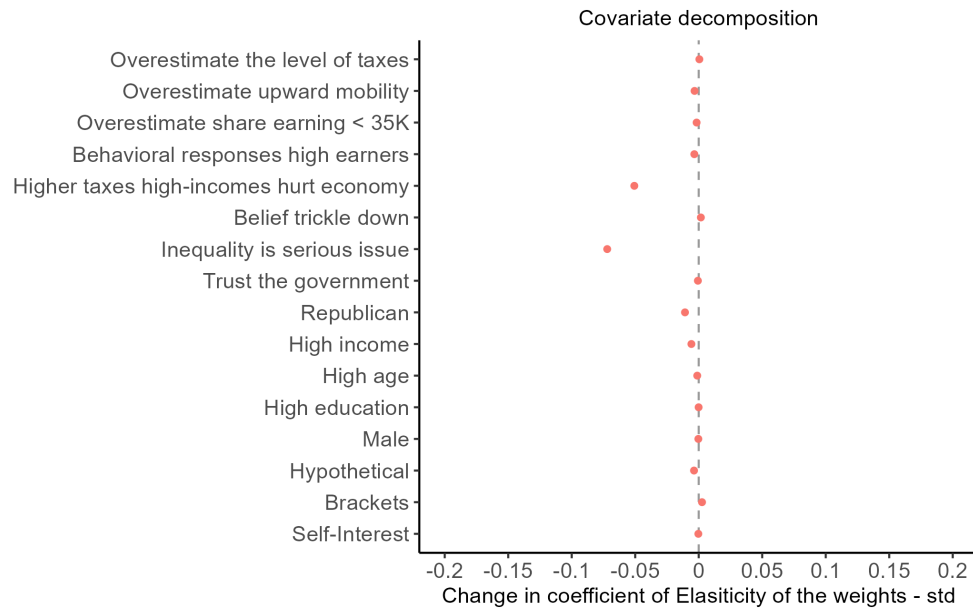
(b) Covariate Decomposition

Figure 5: Predictors of Support for Redistribution and Decomposition of Weights

Notes: Panel (a) presents coefficient estimates. The dependent variable (*Redistribution*) takes values from -2 to +2, where positive (negative) values indicate that income should be further redistributed by taking from higher-income (lower/middle-income) individuals and giving to lower/middle-income (higher-income) individuals. A value of zero indicates that incomes should not be further redistributed. See the main text for an explanation of the explanatory variables. Models s1 and s2 are weighted using sampling weights. Error bars are computed using HC3 standard errors. Panel (b) presents the effect of each variable on the co-efficient estimate of *Elasticity of the weights - std*. The effects are computed using the covariate decomposition procedure proposed by Gelbach (2016). The figures use data from Wave 2.

weights.

To identify which factors are captured by Social Architects' assigned welfare weights, we proceed in two steps. First, we estimate the overall variation in support for redistribution that can be explained by welfare weights using a linear regression. Some of this overall variation may be explained by factors orthogonal to welfare preferences. A factor that predicts Social Architects' support for redistribution via their welfare weights is thus captured by welfare weights. In the second step, we decompose this overall variation into the variation explained by each of the other factors.[36]

The overall variation in support for redistribution that is explained by welfare weights can be estimated by regressing Social Architects' support for redistribution on their standardized elasticity of the weights. The results are presented in Model s1 in Figure 5a. We find that the coefficient estimate of the elasticity of the weights is -0.32.

Some of the overall variation in support for redistribution depicted in Model s1 may be explained by factors orthogonal to welfare preferences. Thus, if we add other variables (capturing other factors) to the regression, the coefficient estimate of the elasticity of the weights would change. Details about the other variables can be found in Appendix Section A. The results are presented in Model s2 in Figure 5a. We find that the coefficient estimate of the elasticity of the weights drops, in absolute value, from -0.32 in Model s1 to -0.17 in Model s2. Thus, the coefficient estimate halves in Model s2 when we account for other factors in the regression, indicating that a part of the overall variation in support for redistribution explained by welfare weights goes through these other factors.

How much of the change in the coefficient estimate of the elasticity of the weights (between Model s1 and s2) is explained by each of the other variables? Figure 5b presents the effect of each variable on the coefficient estimate of the elasticity of the weights. The effects are computed using the covariate decomposition procedure proposed by Gelbach (2016). A negative value for a variable indicates that its inclusion in the regression reduces the absolute value of the coefficient estimate of the elasticity of the weights in Model s1. We find that the change in the coefficient estimate of the elasticity of weights is almost entirely driven by Social Architects' beliefs about the externalities due to inequality and their beliefs about higher taxes on high-income individuals hurting the economy.

Factors that affect Social Architects' support for redistribution via their welfare weights are captured by our measure of welfare weights. Thus, Social Architects' welfare weights partly capture their beliefs about the externalities due to inequality and their beliefs about higher taxes on high-income individuals hurting the economy. This leads to the following

---

[36]Our estimation assumes that the set of concerns captured by support for redistribution is a superset of the set of concerns captured by welfare weights.

result.

Result 5. *Social Architects' welfare weights partly capture their beliefs about the externalities due to inequality and their beliefs about higher taxes on high-income individuals hurting the economy.*

Interestingly, we find that Social Architects' support for redistribution captures many factors orthogonal to welfare preferences that are not captured by their welfare weights. This result suggests that using the experimental measure of welfare weights can be useful in settings where we want to capture variation in welfare preferences rather than factors orthogonal to welfare preferences.

5 Comparing Welfare Weights

In this section, we compare the general population weights to the weights implied by tax and transfer policies and weights used in the optimal policy literature. For these comparisons, we focus on Treatments Loss × 70K, Gain × 70K, and Real. The range of the average elasticity estimates across these three treatments is given by $\nu \in [-0.37, -0.53]$.[37] We also identify the range of the median elasticity estimates across these three treatments. In particular, we compute the weighted median value of ν using Equation (6), weighting the sample using sampling weights. We find that the range of the median progressivity of the welfare weights is given by $\nu \in [-0.30, -0.6]$, which is similar to the range of estimates of the average progressivity.

5.1 Comparing to Weights Implied by Tax and Transfer Policies

How do the welfare weights of the general population compare to the weights implied by the income tax schedule and transfer policies in the U.S.? The weights implied by the tax schedule and transfer policies can be represented as politicians' aggregation of societal welfare weights, potentially influenced by political economy considerations.

We obtain the weights implied by the income tax schedule in the U.S. from Hendren (2020).[38] Hendren (2020) uses the universe of tax returns in 2012, and compute indi-

---

[37]We exclude the remaining treatments for various reasons. Treatments Loss × 500K and Gain × 500K are excluded because these treatments have a limited range of possible values of ν. This limitation is discussed in Section 4.6. We exclude Treatment Hypothetical because we find that the elicited welfare weights are different in Treatments Hypothetical and Real; we defer to Treatment Real since the presence of real stakes is likely to lead to more reliable welfare weights. We exclude Treatment No Self-Interest because these treatments presented the income brackets of the Recipients rather than the exact incomes. This treatment does not control for Social Architects' beliefs about Recipients' incomes. For example, Social Architects could believe that the Recipient earning in the income bracket "$375,000 and above" earns $375,000 or $2,000,000. The estimated elasticity in this treatment can be sensitive to the assumptions about Social Architects' beliefs about Recipients' incomes. Finally, Treatment Self-Interest is excluded because we follow the tradition in welfare economics of exploring people's views unconfounded by self-interest motives.

[38]The weights implied by the tax schedule are obtained using the optimal income tax formula. The opti-

viduals' tax liabilities based on ordinary income taxes, alternative minimum tax (AMT), earned income tax credits (EITC), state and local taxes, and Medicare.

The weights implied by transfer policies in the U.S. are derived based on the framework outlined in Hendren and Sprung-Keyser (2020). In this framework, the desirability of a policy can be assessed by its Marginal Value of Public Funds (MVPF). The MVPF indicates the welfare accrued to the beneficiaries of a policy from \$1 of government spending. There exists an inverse relationship between the MVPF of a policy and the welfare weights assigned to its beneficiaries. Intuitively, a policy with a low MVPF may be deemed desirable if a high welfare weight is placed on its beneficiaries.[39] Hendren and Sprung-Keyser (2020) provide estimates of the MVPF of various policies. We use their estimates to compute the welfare weights implied by these policies. We focus on taxes, cash transfers, and in-kind transfers, as they are the policies most similar to those in our study.[40] Additionally, we restrict the sample to policies with a positive MVPF, as our theoretical framework and experimental design cannot accommodate negative welfare weights.[41]

To compare these different sets of welfare weights, we assume a parametric form for the welfare weights, as done in the previous sections, in which the welfare weights depend on the incomes of the Recipients. In particular, we compute the elasticity of the welfare weights with respect to Recipients' incomes. Figure 6 plots the welfare weights interpolated using the function y^ν against the disposable income distribution. We find that the elasticity of the weights implied by the income tax schedule is -0.10 and by transfer policies is -0.27. The average progressivity of general population weights ($\nu \in [-0.37, -0.53]$) is 3.7 to 5.3 times more progressive than the progressivity of the weights

mal income tax formula calculates the optimal marginal tax rates as a function of welfare weights and other relevant objects, such as the elasticity of taxable income. The formula can be inverted to obtain the "inverse-optimum" tax formula. The inverse-optimum welfare weights that make the current tax schedule optimal are derived from the inverse-optimum tax formula and the current marginal tax rates. We present the welfare weights implied by the income tax schedule at each quintile of the income distribution in Figure A8 in Appendix Section C.1.

[39]Consider a policy that affects Recipients with incomes near z^*. The Marginal Value of Public Funds (MVPF) of a policy is defined as the Recipients' willingness to pay for the policy (s^*) divided by the net cost (c) accrued from the policy to the government. If the government aims to achieve s^* through adjustments to the tax schedule instead of implementing the policy, the cost to the government would be $s^*g(z^*)$, where $g(z^*)$ is the marginal value of an additional dollar of consumption (welfare weight). It would be cheaper for the government to achieve s^* through the policy than through adjustments to the tax schedule if and only if $c \leq s^*g(z^*)$. Rewriting this expression yields the following equation: $MVPF = s^*/c \geq 1/g(z^*)$.

[40]These policies are similar to those in our study because they are only conditioned on Recipients' incomes and they are guided by redistributive motives.

[41]We list the selected policies in Table A6 in Appendix Section C.2. We present the welfare weights implied by the selected policies against the average income of the beneficiaries of the policies in Figure A9 in Appendix Section C.1.

implied by the income tax schedule and 1.4 to 2 times more progressive than the weights implied by transfer policies.

The above results suggest that the weights implied by the tax schedule are less progressive than the general population weights. The general population weights were elicited based on Recipients' disposable incomes. If the general population weights were elicited based on Recipients' pre-tax income, they would likely be even more progressive. This suggests that the observed gap between the general population weights and the weights implied by the tax schedule is a lower bound. We now have the following result.

Result 6. *The general population welfare weights are 3.7 to 5.3 times more progressive than the weights implied by the income tax schedule in the U.S. and 1.4 to 2 times more progressive than the weights implied by transfer policies in the U.S.*
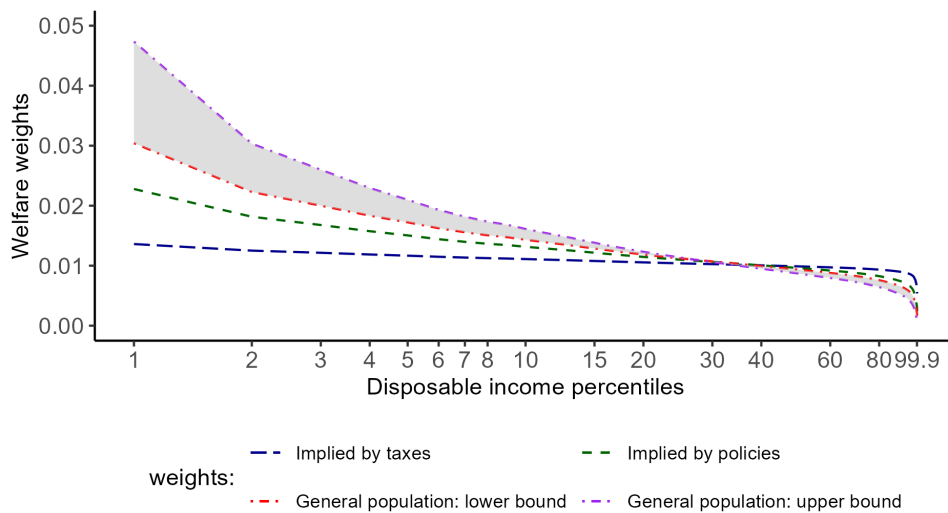

Figure 6: General Population Weights and Weights Implied by Tax and Transfer Policies

Notes: The figure plots welfare weights against percentiles of the disposable income distribution. The x-axis has a natural log spacing. We use the function y^ν to interpolate the welfare weights for the disposable income (y) distribution and then re-normalize the welfare weights such that they sum to 1. The figure plots the re-normalized weights implied by the income tax schedule computed by Hendren (2020) ($\nu = -0.1$), weights implied by transfer policies computed by Hendren and Sprung-Keyser (2020) ($\nu = -0.26$), lower bound of the general population weights ($\nu = -0.37$), and upper bound of the general population weights ($\nu = -0.53$). Details on the construction of the figure can be found in Appendix Section E.2.

Decomposing the Gap

Why are the general population weights more progressive than the weights implied by the income tax schedule? We represent politicians' actions as an aggregation of societal

welfare weights and explore alternative assumptions about aggregation that can explain the observed gap between the two sets of weights. This exercise assumes that the welfare weights elicited in the experiment are the "true" welfare weights of the general population. We discuss alternative explanations below.

First, we compare the weights implied by the tax schedule to the median general population weights. Under the median voter theorem, politicians' aggregation of welfare weights corresponds, under some assumptions, to the median welfare weights. We compute the weighted median value of ν using Equation (6), weighting the general population sample using sampling weights. We find that the estimated value of ν is -0.45, suggesting that the general population weights remain more progressive than the weights implied by the tax schedule.

Second, we consider alternative assumptions about the group of individuals that the politicians consider when aggregating societal welfare weights. In particular, politicians may be considering only the voting population, not the entire population. We find that the weighted median value of ν, weighting the general population sample using sampling weights and the likelihood of being a registered voter, is -0.47.[42] This result suggests that accounting for the voting population does not close the gap between the general population weights and the weights implied by the tax schedule.

Finally, we explore the hypothesis of "elite capture," which posits that politicians overweight the interests of high-income individuals when aggregating societal preferences. We observe an upward spike in the welfare weights implied by the tax schedule for the two highest percentile in Figure A8 in Appendix Section C.1. Does this mean that politicians overweight the welfare weights of high-income individuals? To test the hypothesis of elite capture, we compute the weighted average progressivity of the welfare weights of the general population, weighting the sample by sampling weights and various "aggregation" weights. We assign an aggregation weight of x to all individuals earning less than \$100,000 and an aggregation weight of $1 - x$ to all individuals earning more than \$100,000. We find that the weighted average of ν decreases with $1 - x$ since higher-income individuals in our sample have less progressive welfare weights. The weighted median estimate of ν drops to -0.38 when the aggregation weight assigned to higher-income Social Architects is nine times higher than the aggregation weight assigned to lower-income Social Architects. Thus, the gap between the general population weights and the weights implied by the income tax schedule is partly driven by politicians overweighting the welfare weights of high-income individuals. This explanation is consistent with the evidence in the literature suggesting that implemented policies are often more likely to reflect the

---

[42]Details on the analysis can be found in Appendix Section E.3.

interests of high-income individuals (e.g., Gilens and Page, 2014; Karabarbounis, 2011).

Several other factors can lead to a gap between the general population weights and the weights implied by the tax schedule. First, the gap could arise if the welfare weights elicited in our experiment are not the welfare weights that politicians aggregate. For example, politicians may be aggregating the weights of individuals guided, at least partly, by self-interest motives.[43] In addition, politicians may be aggregating the welfare weights of individuals who may be guided by factors that matter in the real world but which are not captured (intentionally) by the welfare weights in our experiment, such as trust in government. Finally, the gap could arise if politicians' beliefs about the behavioral responses to taxation, captured by the elasticity of taxable income, differ from the estimates used by Hendren (2020) to compute the weights implied by the tax schedule.

5.2 Comparing to Weights Used in the Literature

How do the general population welfare weights compare to the welfare weights used in the optimal policy literature? While the literature uses different estimates of welfare weights, one frequently used estimate is inversely proportional to Recipients' disposable incomes or consumption (e.g., Saez, 2001). We refer to these as "inverse-consumption" weights. These welfare weights can be derived by assuming a utilitarian social welfare function and (i) log utilities or (ii) constant relative risk aversion (CRRA) in consumption utilities with a coefficient of relative risk aversion given by $\gamma = 1$.[44]

Inverse-consumption weights can be characterized by the function y^ν with $\nu = -1$. We find that these inverse-consumption weights are 1.8 to 3.3 times more progressive than the general population weights ($\nu \in [-0.37, -0.53]$). We plot the inverse-consumption weights and the general population weights in Figure A10 in Appendix Section C.2. We have the following result.

Result 7. *Welfare weights that are inversely proportional to Recipients' consumption are about 1.8 to 3.3 times more progressive than the general population weights.*

There is a mapping from the welfare weights obtained assuming CRRA utilities with

---

[43]In Section 4.6, we find that Social Architects' weights are more progressive in Treatment Self-Interest relative to Treatment No Self-Interest. This suggests that allowing for self-interest motives to affect welfare weights would widen the gap between the general population weights and weights implied by the tax schedule.

[44]CRRA utilities are given by $u(c, l) = u(c) + v(l) = \frac{c^{1-\gamma}-1}{1-\gamma} - v(l)$, where c represents consumption, v represents the disutility of work (which depends on labor effort l), and γ is the coefficient of relative risk aversion. Log utilities are given by $u(c, l) = log(c) - log(v(l))$. Taking the first-order condition of the utilitarian social welfare function $W = \sum u(c, l)$ results in $\Delta W = \sum u'(c)$. For CRRA utilities, we have $\Delta W = c^{-\gamma} \Delta c$, where $u'(c) = c^{-\gamma}$ characterizes the welfare weights. Thus, inverse-consumption weights correspond to $\gamma = 1$. For log utilities, we have $\Delta W = \frac{1}{c} \Delta c$ where $u'(c) = \frac{1}{c}$ characterizes the welfare weights.

a coefficient of relative risk aversion γ to the parametric form y^ν, given by $\gamma = -\nu$. Our results suggest that the optimal policy literature could use CRRA utilities with $\gamma \in [0.37, 0.53]$ to generate policies that would be socially acceptable in the U.S.

5.3   Calibrating Optimal Labor Income Taxes

We explore the implications of the estimated welfare weights for the optimal non-linear labor income taxes in the U.S. We use the optimal tax formula derived in Saez (2001). The formula is the solution to the planner's problem of maximizing social welfare, given the constraints on government revenue. It provides the optimal marginal tax rates for different incomes as a function of (i) the shape of the income distribution, (ii) welfare weights, and (iii) the elasticity of taxable income. We use the actual income distribution in the U.S. to identify the shape of the income distribution. The estimates of the elasticity of taxable incomes are taken from the literature. Finally, we use the different estimates of welfare weights computed in the previous section. The progressivity of the optimal marginal tax rates is increasing with the progressivity of the welfare weights. Our calibration is based on Mankiw et al. (2009) and Støstad and Cowell (2022). Details on the calibration of the tax formula are provided in Appendix Section E.4.

Figure 7 plots the optimal marginal tax rates for different estimates of welfare weights against the labor income distribution. The figure also plots the actual marginal tax rates.[45]

The optimal tax rates exhibit a U-shape pattern, mirroring the pattern commonly observed in the literature. The optimal marginal tax rate calibrated using inverse-consumption welfare weights ($\nu = -0.1$), which are the most progressive among the set we consider, is depicted in the top line. The average tax rate in this case is 62%. This is similar to the average rate of 59% found in Saez (2001).

The average optimal marginal tax rate is 45% when based on the lower bound of the general population weights ($\nu = -0.37$) and 51% when based on the upper bound ($\nu = -0.53$). We find that the optimal rates based on the general population weights are lower than those obtained using inverse-consumption weights. The difference in the average rates is 11-17 percentage points.

The average rate based on the weights implied by transfer policies ($\nu = -0.26$), the second least progressive weights among the set, is 40%. The welfare weights implied by the income tax schedule ($\nu = -0.1$), which are the least progressive among the set we consider, lead to the least progressive optimal marginal tax rates. The average rate in this case is 24%. The optimal marginal tax rates based on the general population weights

---

[45]We compute the ordinary income taxes based on the marginal tax rates for single filers in 2019. Following Hendren (2020), we add a 5% state tax, a 2.9% tax rate for Medicare, and a 2.4% sales tax rate

are higher than those obtained using the weights implied by tax and transfer policies. In particular, the average rate based on the general population weights is 5-11 percentage points higher than that obtained using the weights implied by transfer policies and 21-27 percentage points higher than that obtained using the weights implied by the income tax schedule.

The actual marginal tax rates are depicted by the solid black line. The optimal marginal tax rates calibrated using the general population weights are higher than the actual marginal tax rates, except for incomes between $150,000 and $200,000. The average actual marginal taxes are 29%. The average rates based on the general population weights are 16-22 percentage points higher than the actual rates.
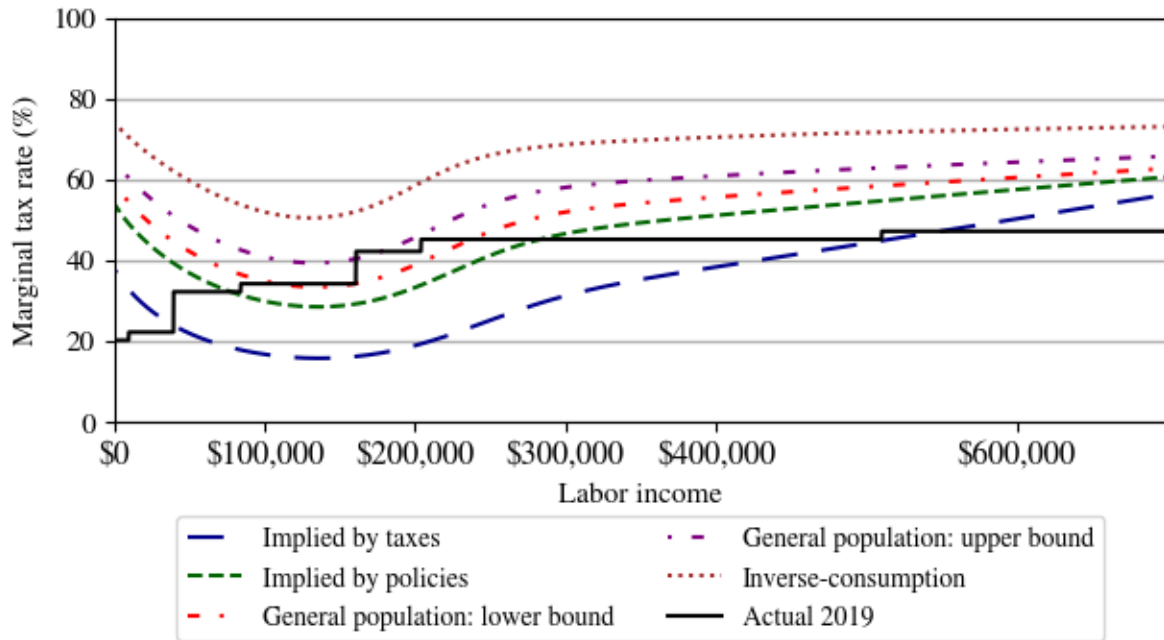

Figure 7: Marginal Tax Rates

Notes: The figure plots marginal tax rates against the labor income distribution. The figure plots the optimal marginal tax rates obtained using the weights implied by the income tax schedule computed by Hendren (2020) ($v = -0.1$), weights implied by transfer computed by Hendren and Sprung-Keyser (2020) ($v = -0.26$), lower bound of the general population weights ($v = -0.37$), upper bound of the general population weights ($v = -0.53$), and inverse-consumption weights ($v = -1$). The figure also plots the actual marginal tax rates. See Appendix Section E.4 for details.

Overall, a key result with policy implications is that the optimal marginal tax rates calibrated using the general population weights are lower than those calibrated using inverse-consumption weights but higher than the current income tax schedule. We now have the following result.

Result 8. *The optimal marginal tax rates calibrated using the general population weights are lower than those calibrated using inverse-consumption weights but higher than the current income tax rates.*

6   Discussion

We elicit the welfare weights assigned by the general population of the U.S. We conducted two real-stakes online experiments with large samples of the general population in the U.S. (N ≈ 4000). In our experiment, participants in the role of "Social Architects" assign welfare weights to participants in the role of "Recipients." In our framework, a reform is conditioned solely on Recipients' incomes. In the experiment, Social Architects assign welfare weights based on Recipients' incomes. A Social Architect's assigned welfare weights can be guided by various underlying ideals, such as equality of opportunity or utilitarianism. However, knowing a Social Architect's welfare weights in the experiment is sufficient to learn their assessment of the welfare implications of a reform and, by extension, its desirability.

We find that the general population weights are progressive on average. This indicates that, on average, the general population wants additional redistribution beyond that achieved by the current tax and transfer system. The general population weights are more progressive than the weights implied by the current tax and transfer policies in the U.S. Our results are consistent with the explanation that politicians, aggregating societal welfare weights, overweight high-income individuals who assign less progressive weights. The general population weights are less progressive than the weights frequently used in the optimal policy literature, which are inversely proportional to Recipients' incomes. We provide estimates of the general population weights that the optimal policy literature can use to obtain socially acceptable policies in the U.S. Finally, we explore the implications of the general population weights for the optimal non-linear income taxes in the U.S. We find that the optimal marginal tax rates obtained using the general population weights are lower than the rates obtained using the weights frequently used in the optimal policy literature but higher than the current rates in the U.S.

6.1   Sufficient-Statistics Approach

Our sufficient-statistics approach to eliciting welfare weights and the "small-reform" approach to taxation used by Saez and Stantcheva (2016) has a few limitations.

First, the welfare weights assigned by the Social Architects may not be normatively appealing if they have inaccurate beliefs at various margins. However, there are two reasons why the presence of inaccurate beliefs may not be problematic. First, Social Architects may have misperceptions at various margins, but their assigned welfare weights may

43

not be sensitive to their misperceptions. We show that Social Architects' welfare weights are not sensitive to misperceptions about income mobility, the share of low-income individuals, and the level of taxes paid by individuals. Nevertheless, we cannot rule out the possibility that Social Architects' misperceptions at other margins affect their welfare weights. Second, the average welfare weights of Social Architects may be unaffected by misperceptions if the misperceptions across Social Architects sum to zero.

Second, the welfare weights estimated in our paper cannot be used to evaluate non-marginal ("large") reforms.[46] For non-marginal reforms, the marginal value of the first dollar may not be the same as the marginal value of the last dollar. For example, a Social Architect with progressive welfare weights may find a marginal budget-neutral reform desirable but may find a non-marginal reform that completely equalizes incomes as undesirable. We assume a parametric form for the welfare weights, in which the welfare weights depend on Recipients' disposable incomes, which allows the welfare weights to be applicable to evaluate large reforms.

Third, our approach assumes that tax reforms have no general equilibrium effects or spillover effects. Spillover effects could imply that a reform affects individuals who were not the direct target of the reform. Evaluating a reform that involves these effects can be challenging since the elicited welfare weights may not account for such effects.

Fourth, the welfare weights may not be applicable across time and countries. There is evidence in the literature suggesting that people's support for redistribution may differ over time (Fisman et al., 2015) and across countries (e.g., Almås et al., 2020; Falk et al., 2018). Future work can test whether welfare weights differ across time and countries.

6.2 Aggregation of Welfare Weights

A key question in our setting relates to the aggregation of Social Architects' welfare weights. In our paper, we aggregate Social Architects' welfare weights by computing the average or the median progressivity of the welfare weights. Assuming that a democratic government aggregates societal welfare weights, the aggregate weights under the median voter theorem correspond to the median weights. We then use these aggregate welfare weights to evaluate policies. There are two potential concerns regarding aggregation.

First, the median progressivity of the weights may not be the most normatively appealing aggregation. Exploring different methods of aggregation is an important area for future research. One avenue is to take a positive approach and explore how the general population would aggregate welfare weights. A similar approach was taken by Ambuehl and Bernheim (2021) in the context of ordinal preference aggregation.

---

[46]It is worth noting that alternative approaches, such as using the inverse-optimum weights implied by policies, also cannot be used to evaluate non-marginal reforms (Hendren, 2020).

Second, it is unclear whether policy formulas should use the aggregate welfare weights of the Social Architects or the individual welfare weights. If individual weights are used, the resulting policies would then need to be aggregated across all Social Architects.

6.3 Future Research

There are several avenues for future research. First, future work can explore whether people's welfare preferences depend on the policy domain. For example, people may want to assign different weights in the healthcare domain than in the education domain. Our portable experimental design can be used to estimate welfare weights in other domains. Second, future research can elicit welfare weights for policies that are conditioned on Recipients' incomes as well as their personal characteristics ("tags").

References

Alesina, A., Stantcheva, S. and Teso, E. (2018), 'Intergenerational mobility and preferences for redistribution', *American Economic Review* **108**(2), 521–554.

Allcott, H., Lockwood, B. B. and Taubinsky, D. (2019), 'Regressive sin taxes, with an application to the optimal soda tax', *The Quarterly Journal of Economics* **134**(3), 1557–1626.

Almås, I., Cappelen, A. and Tungodden, B. (2020), 'Cutthroat Capitalism versus Cuddly Socialism: Are Americans more Meritocratic and Efficiency-seeking than Scandinavians?', *Journal of Political Economy* **128**, 1753–1788.

Ambuehl, S. and Bernheim, B. D. (2021), Interpreting the will of the people: a positive analysis of ordinal preference aggregation, Technical report, National Bureau of Economic Research.

Bourguignon, F. and Spadaro, A. (2012), 'Tax–benefit revealed social preferences', *The Journal of Economic Inequality* **10**(1), 75–108.

Bursztyn, L., Egorov, G., Haaland, I., Rao, A. and Roth, C. (2023), 'Justifying dissent', *The Quarterly Journal of Economics* **138**(3), 1403–1451.

Chen, D. L., Schonger, M. and Wickens, C. (2016), 'otree—an open-source platform for laboratory, online, and field experiments', *Journal of Behavioral and Experimental Finance* **9**, 88–97.

Cohn, A., Jessen, L. J., Klasnja, M. and Smeets, P. (2019), Why do the rich oppose redistribution? An experiment with america's top 5%. Working paper.

Diamond, P. and Saez, E. (2011), 'The case for a progressive tax: from basic research to policy recommendations', *Journal of Economic Perspectives* **25**(4), 165–90.

Drenik, A. and Perez-Truglia, R. (2018), 'Sympathy for the diligent and the demand for workfare', *Journal of Economic Behavior & Organization* **153**, 77–102.

Enke, B., Graeber, T. and Oprea, R. (2023), 'Confidence, self-selection, and bias in the aggregate', *American Economic Review* **113**(7), 1933–1966.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018), 'Global evidence on economic preferences', *The Quarterly Journal of Economics* **133**(4), 1645–1692.

Fisman, R., Jakiela, P. and Kariv, S. (2015), 'How did distributional preferences change during the great recession?', *Journal of Public Economics* **128**, 84–95.

Fisman, R., Jakiela, P. and Kariv, S. (2017), 'Distributional preferences and political behavior', *Journal of Public Economics* **155**, 1–10.

Fleurbaey, M. and Maniquet, F. (2006), 'Fair income tax', *The Review of Economic Studies* **73**(1), 55–83.

Gelbach, J. B. (2016), 'When do covariates matter? and which ones, and how much?', *Journal of Labor Economics* **34**(2), 509–543.

Gilens, M. and Page, B. I. (2014), 'Testing theories of american politics: Elites, interest groups, and average citizens', *Perspectives on politics* **12**(3), 564–581.

Haaland, I. and Roth, C. (2023), 'Beliefs about racial discrimination and support for pro-black policies', *Review of Economics and Statistics* **105**(1), 40–53.

Haaland, I., Roth, C. and Wohlfart, J. (2023), 'Designing information provision experiments', *Journal of economic literature* **61**(1), 3–40.

Hendren, N. (2020), 'Measuring economic efficiency using inverse-optimum weights', *Journal of Public Economics* **187**, 104198.

Hendren, N. and Sprung-Keyser, B. (2020), 'A unified welfare analysis of government policies', *The Quarterly Journal of Economics* **135**(3), 1209–1318.

Karabarbounis, L. (2011), 'One dollar, one vote', *The Economic Journal* **121**(553), 621–651.

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', *Journal of Economic Perspectives* **23**(4), 147–174.

Saez, E. (2001), 'Using elasticities to derive optimal income tax rates', *The review of economic studies* **68**(1), 205–229.

Saez, E. (2002), 'Optimal income transfer programs: intensive versus extensive labor supply responses', *The Quarterly Journal of Economics* **117**(3), 1039–1073.

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

Singhal, M. (2008), 'Quantifying preferences for redistribution', *NBER Working Paper, Harvard University, John F. Kennedy School of Government, Cambridge* .

Stantcheva, S. (2016), 'Comment on "positive and normative judgments implicit in us tax policy and the costs of unequal growth and recessions" by benjamin lockwood and matthew weinzierl', *Journal of Monetary Economics* **100**(77), 48–52.

Stantcheva, S. (2020), 'Understanding economic policies: What do people know and learn', *Unpublished Manuscript, Harvard University* pp. 2309–2369.

Støstad, M. N. and Cowell, F. (2022), 'Inequality as an externality: Consequences for tax design', *Available at SSRN 4185685* .

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

Weinzierl, M. (2018), 'Revisiting the classical view of benefit-based taxation', *The Economic Journal* **128**(612), F37–F64.

Zoutman, F. T., Jacobs, B. and Jongen, E. L. (2013), 'Optimal redistributive taxes and redistributive preferences in the netherlands', *Working paper* .

Online Appendix

Who Should Get Money? Estimating Welfare Weights in the U.S.

Contents

A Variable Definitions

Misperceptions

We elicit Social Architects' perceptions about the level of taxes paid by individuals, the share of individuals with incomes below \$35,000, and upward mobility.

Overestimate the level of taxes: We ask Social Architects four questions designed to elicit their perceptions about the level of taxes paid by individuals in society. In particular, Social Architects are asked about their beliefs regarding (i) the share of households in the top tax bracket, (ii) the average tax rate of those in the top tax bracket, (iii) the share of households who pay no taxes, and (iv) the average tax rate of households with the median income. We focus on perceptions along these four dimensions because they were the most predictive of people's support for redistribution in Stantcheva (2021). Social Architects can select a number from 0 to 100 using a slider for each of the four questions. We identify misperceptions in each of the four variables as follows.

- *Gap in top-taxes = Beliefs about top-taxes - 32.7*

- *Gap in top-share = Beliefs about top-share - 0.73*

- *Gap in non-filers = 44 - Beliefs about non-filers*

- *Gap in median-taxes = Beliefs about median-taxes - 13*

We take the actual values from Stantcheva (2021). We orient the gap in non-filers such that a lower gap in non-filers corresponds to an overestimation in the level of taxes paid. We standardize each of the four misperception variables such that they have a mean of 0 and a standard deviation of 1. Then, we create an index by taking the equally weighted average of the four standardized variables and standardizing the resulting variable.

Overestimate share earning $< 35K$: We elicit Social Architects' beliefs about the share of households earning less than \$35,000. This income level corresponds to the income of the Recipient with the second lowest income in our experiment. Social Architects can select a number from 0 to 100 using a slider. We identify Architects' misperceptions by subtracting the actual value (29) from their responses. We obtain the actual value by looking at the share of individuals whose disposable income is below \$35,000 in the data obtained from Piketty et al. (2018) (variable *diinc*). Finally, we standardize the misperceptions.

Perceptions about upward mobility: We ask Social Architects about their perceived likelihood of a child with parents in the first quintile of the income distribution growing up to be in the highest quintile as an adult. Social Architects can select a number from 0 to 100 using a slider. We identify Architects' misperceptions by subtracting the actual value

(7.8) from their responses. We obtain the actual value from Alesina et al. (2018). Finally, we standardize the misperceptions.

Views about Taxes and Government

We ask Social Architects several questions that elicit their views about the tax system and their trust in government, each capturing a unique mechanism that may explain people's support for redistribution. These questions are drawn from Stantcheva (2020). The prompt for the question on inequality being a serious issue is taken from Lobeck and Støstad Nyborg (2022).

Behavioral responses high earners: We ask Social Architects about their beliefs regarding the extent to which taxing high-income earners would encourage them to work less. The indicator variable *Behavioral responses high earners* takes a value of 1 if a Social Architect indicates "A moderate amount," "A lot," or "A great deal," and a value of 0 if the respondent indicates "A little," or "None at all."

Higher taxes high-incomes hurt economy: We elicit Social Architects' beliefs about whether taxing high-income earners would hurt the economy. The indicator variable *Higher taxes high-incomes hurt economy* takes a value of 1 if a Social Architect indicates that taxing high-income earners would "Hurt economic activity in the U.S." and a value of 0 if the Social Architect indicates "Not have an effect on economic activity in the U.S." or "Help economic activity in the U.S."

Belief trickle down: Social Architects are asked whether the lower class and working class would win or lose if taxes on high-income earners were cut. Social Architects who believe in trickle-down economics would believe that if taxes on high-income earners were cut, the lower class and working class would mostly win. *Belief trickle down* is an indicator variable taking a value of 1 if a Social Architect indicates that the lower class and working class would "Mostly win" if taxes on high-income earners were cut and a value of 0 if the respondent indicates that they would "Mostly lose" or "Neither lose nor win."

Inequality is a serious issue: We ask Social Architects their beliefs about inequality being a serious issue, considering that inequality can have externalities on crime, trust, corruption, and social unrest. The indicator variable *Inequality is a serious issue* takes a value of 1 if a Social Architect indicates that inequality is "A serious issue" or "A very serious issue" and a value of 0 if a Social Architect indicates "An issue," "A small issue," or "Not an issue at all."

Trust the government: We ask Social Architects how much of the time they can trust the federal government to do what is right. The indicator variable *Trust the government* takes a value of 1 for the responses "Most of the time" or "Always" and a value of 0 for the responses "Only some times" or "Never."

B Pre-registration

The experimental design, data collection, and analyses were pre-registered.[47] We do not discuss minor deviations from pre-registration in the instructions. We report three deviations from the pre-registration in data collection. First, we received slightly fewer participants than the number we pre-registered in Wave 1. Second, in Wave 1, we relaxed the quotas towards the end of the study to reach our target sample size faster. Third, we excluded participants with multiple survey responses (6 participants from Wave 1 and 7 participants from Wave 2). We did not pre-register this sample restriction as we did not expect participants to take the survey multiple times. Finally, we report some deviations from the pre-registration in the analyses.

1. We construct the sampling weights based on the population estimates in Table 4. These estimates differ by a few percentage points from the incorrect estimates in the pre-registration document.

2. We control for the income of the Social Architects using the variable *High Income*, which is an indicator variable that takes a value of 1 if a Social Architect's income is above the median and 0 otherwise. We pre-registered using the log of the income and the log of the income squared as controls before Wave 1. We pre-registered using *High Income* as a control before Wave 2. We deviate because the variable *High Income* is easier to interpret.

3. We estimate Social Architects' individual-level elasticity of the welfare weights with respect to Recipients' incomes using OLS regressions (by minimizing the sum of squared errors). We pre-registered estimating the elasticity by minimizing the square root of the mean squared errors. While both methods should provide similar results, OLS regressions are computationally simpler to estimate.

4. In the regressions that use a cross-section structure, we use the elasticity of Social Architects' welfare weights with respect to the incomes of the Recipients as our key variable. Although we had pre-registered using the slope of the weights with respect to the Recipients' index (1 to 7) as our key variable, we realize that this measure does not have an economic interpretation and is not portable to other settings.

5. In the section (Appendix Section D.2) exploring the role of Social Architects' own income on their assigned weights, we present regressions in which the explana-

---

[47]https://doi.org/10.1257/rct.8372-3.2

tory variables are a set of seven dummy variables that indicate if a Social Architect's income is near the income of Recipients 1 through 7, respectively. In our pre-registration, we specified including only one dummy variable that indicates if a Social Architect's income is near a given Recipient. We deviate because our current version helps us explore the role of self-interest separately by incomes.

6. We explore the role of self-interest motives (Table A14 in Appendix Section C) by comparing Treatment Self-Interest and Bracket separately for each of the seven income brackets. We had initially pre-registered an incorrect regression.

7. The actual value used to construct the variable *Overestimate share earning < 35K* is 29. We incorrectly pre-registered the value as 30.

8. The results in Table A7 were not pre-registered. Some regressions in Section 4.7.1 and Section D.4 were not pre-registered.

C Additional Tables and Figures

C.1 Additional Figures

Decision Screen 1/6

Please consider each question carefully because if you are selected, one of your choices will have real consequences for two other persons.

	Person A	Person G
After-tax annual income	$8,000	$500,000

Question 1/4: Please choose your preferred alternative

Person A: +$1000 Person G: -$1000	Person A: +$500 Person G: -$500

Figure A1: Screenshot of a Question Presented to Social Architects in Wave 1

Notes: The figure presents a screenshot of a question presented to Social Architects in Wave 1.

Decision Screen 6/6

Please consider each question carefully because if you are selected, one of your choices will have real consequences.

	Person C	You
After-tax annual income	$53,000 to $85,000	$375,000 and above

Question 2/4: Please choose your preferred alternative:

Person C: +$750 You: -$1250 ○	Person C: +$500 You: -$500 ○

Figure A2: Screenshot of a Question Presented to Social Architects in Wave 2

Notes: The figure presents a screenshot of a question presented to Social Architects in Treatment Self-Interest in Wave 2.
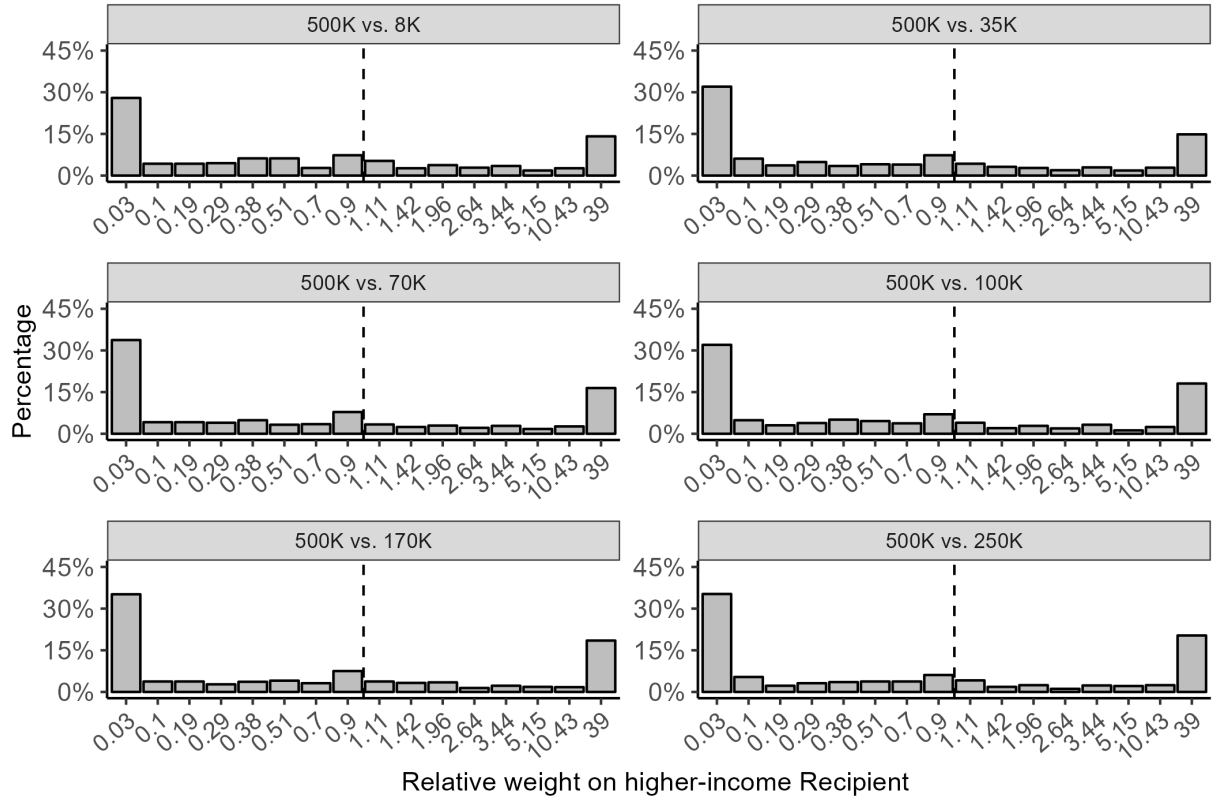
Figure A3: Distribution of $\tilde{g}$ Across Recipient Pairs

Notes: The figure presents the distribution of $\tilde{g}$, the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, across the six pairs of Recipients in the six decision screens. The parameter $\tilde{g}$ is calculated using Equation (4). The sample includes Social Architects in all treatments where the Recipient common across the pairs of Recipients has an income of \$500,000.
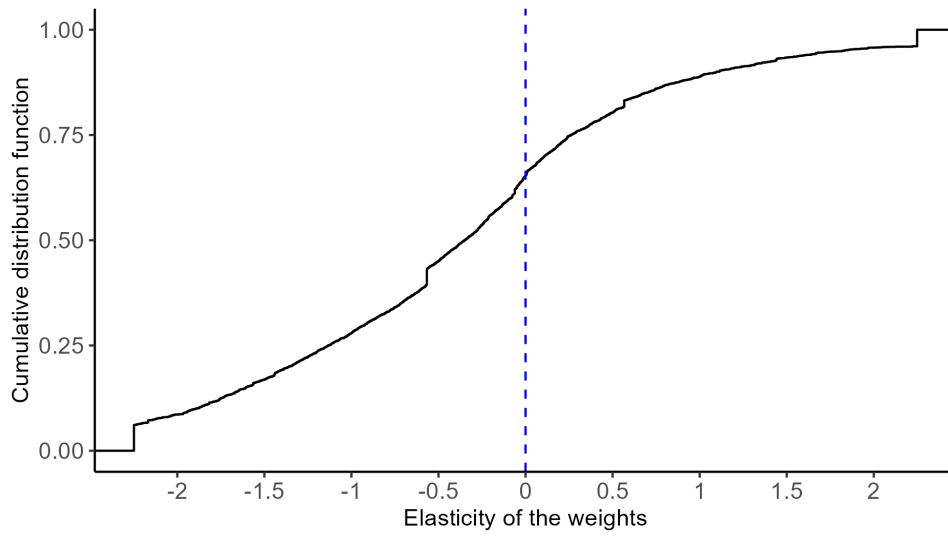
Figure A4: Distribution of the Elasticity of Social Architects' Welfare Weights

Notes: The figure presents the cumulative distribution function (CDF) of the coefficient estimate $\hat{v}$ obtained from the regression $log(g(R_j)) = \beta_0 + vlog(recipient\ income_j) + \epsilon_j$, where $g(R_j)$ is the weight assigned by a given Social Architect to Recipient j, and *recipient income$_j$* is the disposable income of Recipient j.
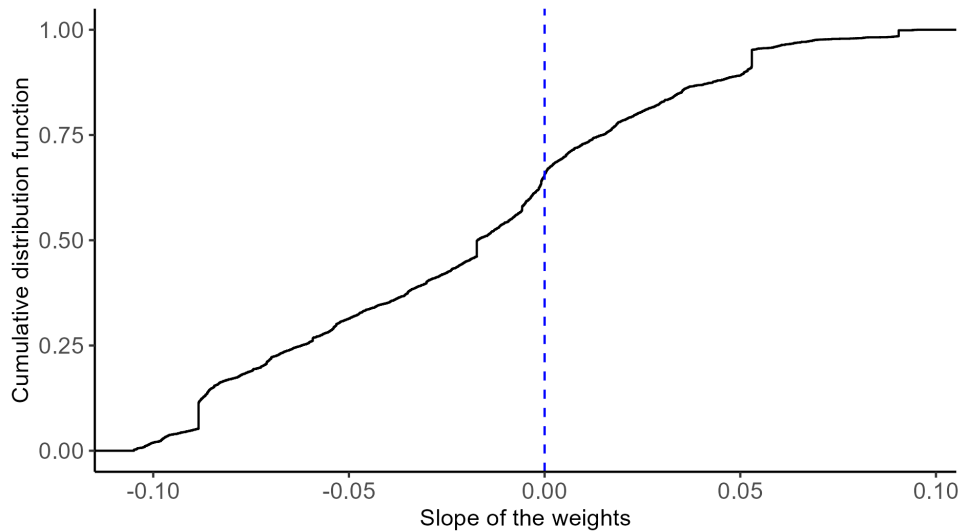
Figure A5: Distribution of the Slope of Social Architects' Welfare Weights

Notes: The figure presents the cumulative distribution function (CDF) of the coefficient estimate $\hat{\beta}_1$ obtained from the regression $g(R_j) = \beta_0 + \beta_1 j + \epsilon_j$, where $g(R_j)$ is the weight assigned by a given Social Architect to Recipient j, and the incomes earned by Recipients 1 through 7 is \$8000 through \$500,000.
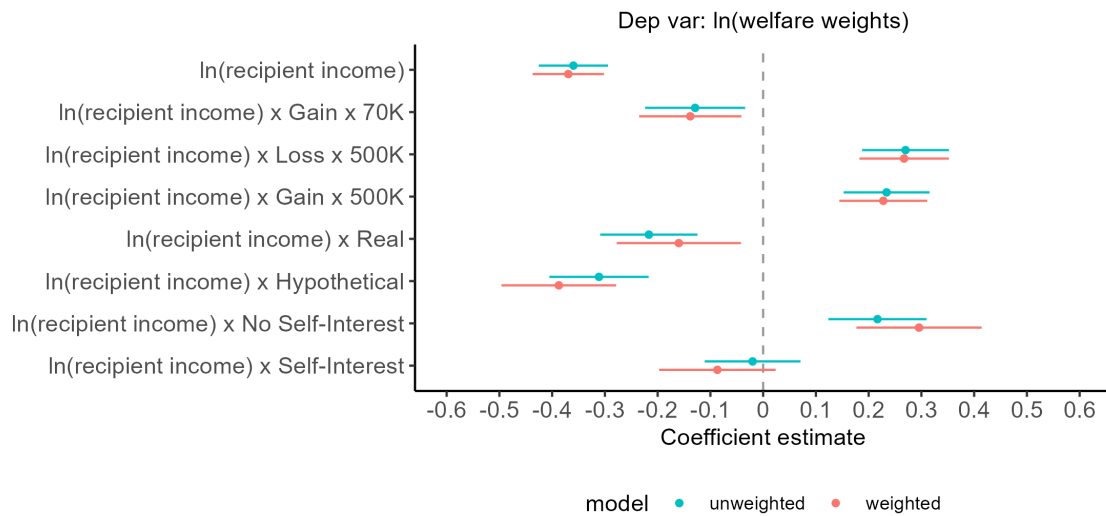
Figure A6: Welfare Weights and Treatments

Notes: The figure presents coefficient estimates. The dependent variable is the log of the welfare weights assigned by Social Architects. The explanatory variables include the log of the incomes of the Recipients, a set of treatment dummies, and the interaction terms of the log of the incomes of the Recipients with the treatment dummies. The base category is Treatment Loss × 70K. We do not present the main effects of the treatment dummies. The regressions weight each treatment using sampling weights. Error bars are computed using HC3 standard errors.
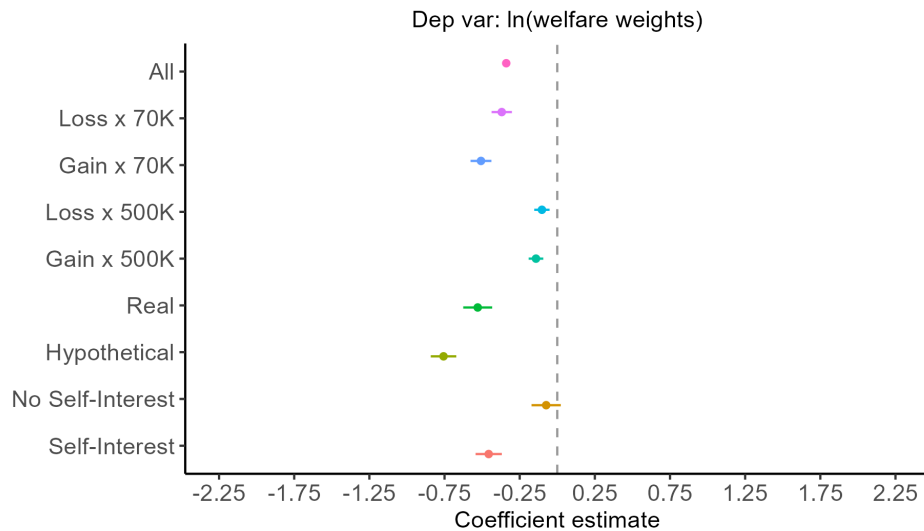
Figure A7: Welfare Weights and Treatments

Notes: The figure presents coefficient estimates. Each dot is a separate regression. The dependent variable is the log of the welfare weights assigned by Social Architects. The explanatory variable is the log of the incomes of the Recipients. The regressions weight each treatment using sampling weights. Error bars are computed using HC3 standard errors.

58
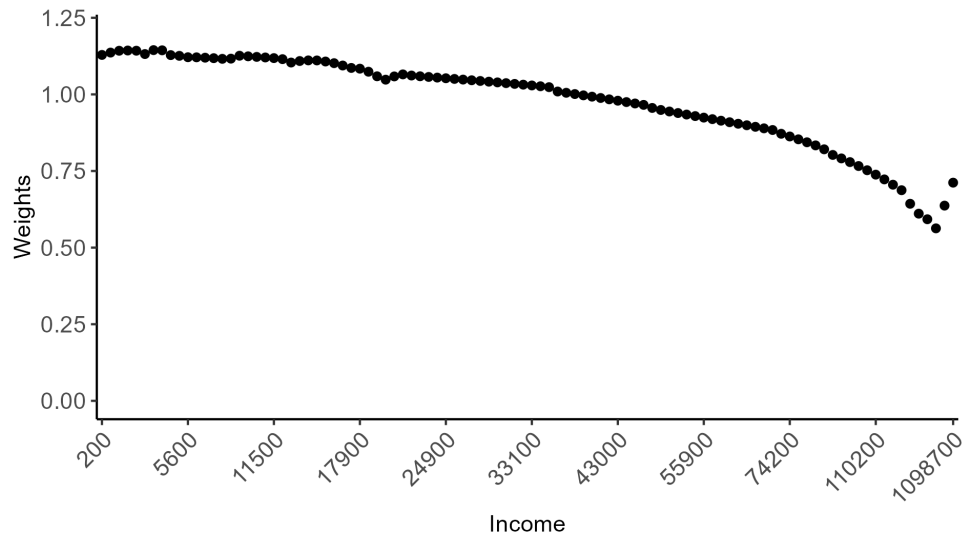
Figure A8: Welfare Weights Implied by the Tax Schedule

Notes: The figure plots the welfare weights implied by the income tax schedule for each quintile of the ordinary income distribution. The welfare weights are computed by Hendren (2020).
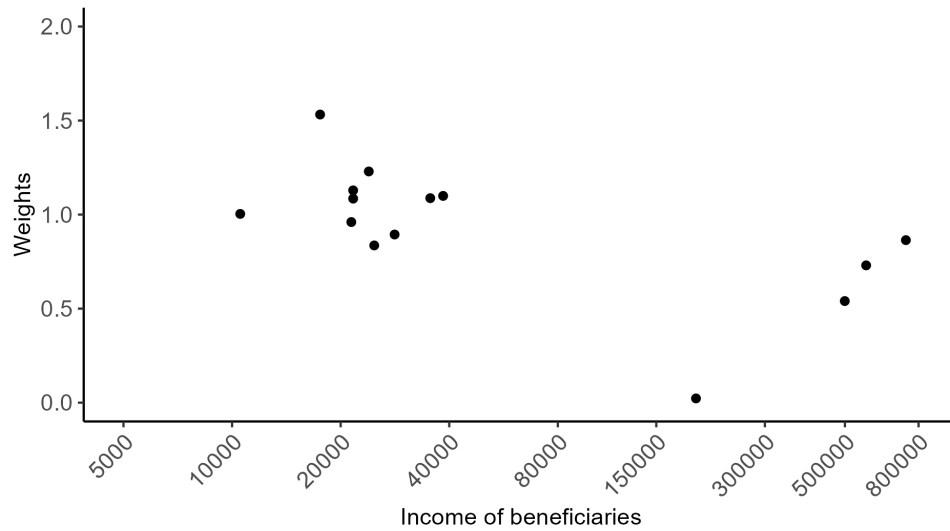
Figure A9: Welfare Weights Implied by Transfer Policies

Notes: The figure plots the welfare weights implied by transfer policies against the average incomes of the beneficiaries of the policies. The x-axis has a log-spacing. The welfare weight of a given policy is the inverse of its marginal value of public funds. The data on the marginal value of public funds of the policies is obtained from Hendren and Sprung-Keyser (2020).
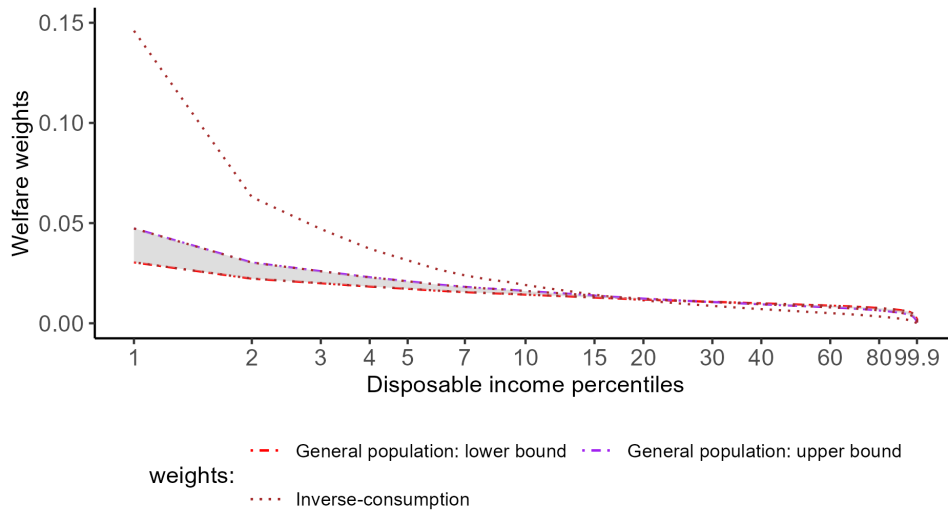
Figure A10: General Population Weights and Inverse-Consumption Weights

Notes: The figure plots welfare weights against percentiles of the disposable income distribution. The x-axis has a natural log spacing. We use the function y^ν to interpolate the welfare weights for the income (y) distribution and then re-normalize the welfare weights to sum to 1. The figure plots the re-normalized inverse-consumption weights ($\nu = -1$), lower bound of the general population weights ($\nu = -0.37$), and upper bound of the general population weights ($\nu = -0.53$). Details on the construction of the figure can be found in Appendix Section E.2.
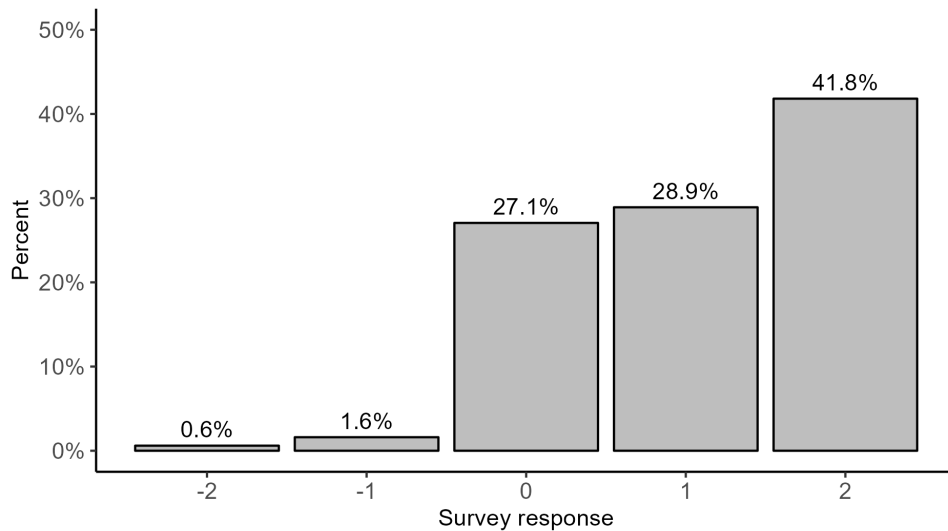
Figure A11: Frequency of Responses to the question *Redistribution*

Notes: This figure illustrates the share of participants selecting each option in the question *Redistribution*. Positive (negative) responses to the question indicate that income should be further redistributed by taking from higher-income (lower/middle-income) individuals and giving to lower/middle-income (higher-income) individuals. A value of zero indicates that incomes should not be further redistributed.

C.2 Additional Tables

Table A1: Summary Statistics by Treatments in Wave 1

	Loss x 500K	Loss x 70K	Gain x 500K	Gain x 70K	p-val
Income: $< 30,000$	0.49	0.54	0.53	0.55	0.00
Income: 30-59,999	0.26	0.26	0.26	0.26	0.53
Income: 60-99,999	0.14	0.12	0.13	0.13	0.00
Income: 100-149,999	0.07	0.03	0.06	0.04	0.02
Income: $> 149,999$	0.04	0.04	0.03	0.03	0.39
Age: 18-34	0.30	0.28	0.29	0.30	0.00
Age: 35-44	0.17	0.16	0.18	0.17	0.12
Age: 45-54	0.18	0.18	0.17	0.17	0.54
Age: 55-64	0.17	0.19	0.16	0.17	0.92
Age: > 64	0.18	0.19	0.20	0.19	0.00
Edu: Up to Highschool	0.44	0.48	0.44	0.46	0.00
Edu: Some college	0.19	0.20	0.20	0.21	0.91
Edu: Bachelor or Associate	0.26	0.23	0.25	0.22	0.00
Edu: Masters or above	0.11	0.10	0.10	0.11	0.00
Region: West	0.19	0.24	0.22	0.20	0.57
Region: North-east	0.17	0.17	0.18	0.19	0.72
Region: South	0.39	0.39	0.41	0.40	0.70
Region: Mid-west	0.24	0.20	0.20	0.21	0.21
Male	0.44	0.53	0.47	0.43	0.00
Republican	0.29	0.35	0.33	0.32	0.00
Minutes Spent	7.42	7.48	7.79	7.66	0.00

Notes: The table presents the average sample characteristics by the four treatments in Wave 1. The last column indicates the p-value corresponding to the F-statistic when testing if the treatment dummies are jointly significant.

Table A2: Summary Statistics by Treatments in Wave 2

	Real	Hypo	No Self	Self	p-val
Income: $< 30{,}000$	0.37	0.39	0.36	0.39	0.00
Income: 30-59,999	0.28	0.28	0.30	0.28	0.32
Income: 60-99,999	0.23	0.20	0.22	0.21	0.00
Income: 100-149,999	0.08	0.10	0.08	0.07	0.00
Income: $> 149{,}999$	0.04	0.03	0.03	0.04	0.99
Age: 18-34	0.37	0.37	0.39	0.37	0.00
Age: 35-44	0.20	0.21	0.22	0.23	0.02
Age: 45-54	0.14	0.14	0.16	0.16	0.74
Age: 55-64	0.19	0.17	0.13	0.16	0.13
Age: > 64	0.10	0.10	0.10	0.09	0.00
Edu: Up to Highschool	0.13	0.15	0.14	0.15	0.00
Edu: Some college	0.20	0.21	0.20	0.20	0.91
Edu: Bachelor or Associate	0.49	0.47	0.48	0.53	0.00
Edu: Masters or above	0.17	0.17	0.18	0.12	0.00
Region: West	0.17	0.14	0.20	0.18	0.01
Region: North-east	0.18	0.21	0.20	0.20	0.19
Region: South	0.42	0.44	0.40	0.43	0.31
Region: Mid-west	0.22	0.20	0.19	0.18	0.38
Male	0.51	0.49	0.48	0.50	0.53
Republican	0.20	0.21	0.18	0.19	0.00
Minutes Spent	18.63	14.14	14.95	15.05	0.00

Notes: The table presents the average sample characteristics by the four treatments in Wave 2: Real, Hypothetical, No Self-Interest, Self-Interest. The last column indicates the p-value corresponding to the F-statistic when testing if the treatment dummies are jointly significant.

Table A3: Classifying Social Architects

Redistributive Tastes	Interval	Share
Weak	$\nu \in (0, |0.625|]$	46%
Strong	$\nu \in (|0.625|, |2|]$	41%
Extremely Strong	$\nu \in (|2|, \infty)$	13%
Weak Progressive	$\nu \in (-0.625, 0]$	28%
Strong Progressive	$\nu \in [-0.625, -2)$	29%
Extremely Strong Progressive	$\nu \in [-2, -\infty)$	9%
Weak Regressive	$\nu \in (0, 0.625]$	19%
Strong Regressive	$\nu \in (0.625, 2]$	12%
Extremely Strong Regressive	$\nu \in (2, \infty)$	4%

Notes: The table presents the share of Social Architects classified by the strength of the reditributive tastes implied type by their welfare weights. The elasticity of the weights for (ν) for each Social Architect is computed using Equation 6

Table A4: Patterns in the Welfare Weights

Variable	Full	Loss x 70K	Real	p-val
Share always choosing CR	0.05	0.06	0.04	0.089
Share always choosing VR	0.10	0.10	0.09	0.934
Share w/ strictly progressive weights	0.00	0.00	0.00	NA
Share w/ strictly regressive weights	0.00	0.00	0.00	NA
Share w/ weakly progressive weights	0.17	0.14	0.24	0
Share w/ weakly regressive weights	0.07	0.07	0.05	0.204
Share w/o weakly monotonic weights	0.77	0.79	0.71	0.008
Slope of the weights	-0.02	-0.02	-0.03	0.001
Share w/ negative slope	0.66	0.63	0.71	0.006
Elasticity of the weights	-0.35	-0.36	-0.58	0.004
Share w/ negative elasiticity	0.65	0.67	0.72	0.121
Maximum observed elasticity	2.25	2.25	2.25	NA
Minumum observed elasticity	-2.25	-2.25	-2.25	NA

Notes: The table presents the average patterns in the Social Architects' assigned welfare weights in the full sample (second column), in Treatment Loss $\times$ 70K (third column), and in Treatment Real (fourth column). The last column indicates the p-value corresponding to the F-statistic when testing if the pattern indicated in the first column is different, on average, between Treatment Loss $\times$ 70K and Treatment Real. The F-statistic is based on a weighted regression that weights each treatment using sampling weights. *Share always choosing CR (VR)* refers to the share who chooses the Constant Reform (Variable Reform) in every question. Strictly progressive (regressive) weights imply that the assigned weights are strictly decreasing (increasing) with the Recipients' income. Weakly progressive (regressive) weights imply that the assigned weights are weakly decreasing (increasing) with the Recipients' incomes. The slope of the weights is the coefficient estimate $\hat{\beta}_1$ obtained from the regression $g(R_j) = \beta_0 + \beta_1 j + \epsilon_j$, where $g(R_j)$ is the weight assigned by a given Social Architect to Recipient j, and the incomes earned by Recipients 1 through 7 are \$8000 through \$500,000. The elasticity of the weights is the coefficient estimate $\hat{v}$ obtained from the regression $log(g(R_j)) = \beta_0 + v log(\textit{recipient income}_j) + \epsilon_j$, where $g(R_j)$ is the weight assigned by a given Social Architect to Recipient j, and *recipient income$_j$* is the income of Recipient j. *Minimum (Maximum) observed elasticity* refers to the minimum (maximum) values of the estimated elasticity of Social Architects' weights observed in the sample.

Table A5: Predictors of the Elasticity of the Weights

Dependent Variable:	Elasticity of the weights		
Model:	(1)	(2)	(3)
Constant	-0.508***	-0.683***	-0.676***
	(0.064)	(0.064)	(0.070)
Gain x 70K	-0.107	-0.125	-0.101
	(0.086)	(0.079)	(0.086)
Loss x 500K	0.291***	0.284***	0.303***
	(0.069)	(0.062)	(0.069)
Gain x 500K	0.249***	0.253***	0.259***
	(0.068)	(0.061)	(0.068)
Real	-0.193**	-0.165**	-0.151*
	(0.084)	(0.076)	(0.084)
Hypothetical	-0.341***	-0.270***	-0.303***
	(0.085)	(0.078)	(0.084)
No Self-Interest	0.228***	0.262***	0.269***
	(0.082)	(0.076)	(0.083)
Self-Interest	-0.024	0.043	0.023
	(0.081)	(0.075)	(0.082)
Screen order	0.282***	0.290***	0.293***
	(0.037)	(0.034)	(0.037)
Republican		0.281***	0.242***
		(0.042)	(0.044)
High income		0.158***	0.174***
		(0.037)	(0.039)
Male		0.017	0.009
		(0.035)	(0.037)
High education		-0.085**	-0.093**
		(0.040)	(0.042)
High age		0.038	0.035
		(0.035)	(0.036)
Observations	3,957	3,957	3,957
Sampling Weights?	Yes	No	Yes

Notes: The table presents linear regressions. The dependent variable is the Social Architects' elasticity of the weights with respect to Recipients' incomes. *High Income*, *High Age*, and *High Education* are indicators of above-median income, age, and education, respectively. *Republican* takes a value of 1 for Republicans and 0 for Democrats or Independents. *Screen Order* indicates the order of the decision screens presented to Social Architects. The weighted regression weights the sample using sampling weights. The standard errors are robust to heteroskedasticity (HC3).
*p<0.1, **p<0.05, ***p<0.01

Table A6: Policies Chosen from Hendren and Sprung-Keyser (2020)

	Policy	Type	Income
1	Top Tax 2013 Increases from Affordable Care Act	Top Taxes	737932
2	Aid to Families with Dependent Children (Greater Benefit Generosity)	Cash Transfers	38467
3	Aid to Families with Dependent Children (Term Limit Modifications)	Cash Transfers	23935
4	Alaska Permanent Fund Dividend	Cash Transfers	35440
5	Top Taxes, Economic Growth And Tax Relief Reconciliation Act 2001	Top Taxes	572569
6	1986 Earned Income Tax Credit Expansion	Cash Transfers	24781
7	1993 Earned Income Tax Credit Expansion	Cash Transfers	28222
8	Effects of Housing Vouchers on AFDC Families Experiment	Housing Vouchers	4395
9	Housing Vouchers in Chicago	Housing Vouchers	17541
10	Top Tax Rate Increase in Omnibus Budget Reconciliation Act 1993	Top Taxes	499532
11	Paycheck Plus Experiment Providing EITC-benefits to Adults without Dependents	Cash Transfers	10535
12	Supplemental Nutrition Assistance Program Application Assitance	Nutrition	21661
13	Supplemental Nutrition Assistance Program Application Information	Nutrition	21661
14	Supplemental Nutrition Assistance Program Introduction	Nutrition	21408
15	Top Tax Rate Reductions in Tax Reform Act of 1986	Top Taxes	193228

Notes: The table presents information about the policies chosen from Hendren and Sprung-Keyser (2020). The third column describes the type of each policy and the fourth column presents the average income of the Recipients of each policy.

D Additional Analysis

D.1 Welfare Weights and Non-welfarist Ideals

we test whether Social Architects' welfare weights capture "non-welfarist" ideals. In the traditional "welfarist" approach, welfare weights depend on characteristics that enter Recipients' utilities (e.g., Recipients' disability status). On the other hand, in the non-welfarist approach, welfare weights do not enter Recipients' utilities (e.g., Recipients' parental income).

In Wave 2, Social Architects are asked about their beliefs regarding whether high-income individuals deserve and need their current income. Their answer options were as follows: (i) "do <u>not</u> deserve their current income and do <u>not</u> need their current income," (ii) "deserve their current income but do <u>not</u> need their current income," (iii) "do <u>not</u> deserve their current income but need their current income," and (iv) "deserve their current income and need their current income." We also ask them a similar question about low-income individuals. The word "need" captures an important class of welfarist ideals in which welfare weights depend on the needs of the Recipients. The word "deserve" captures all non-welfarist ideals.

We explore whether Social Architects' welfare weights capture welfarist and non-welfarist using a regression. Our dependent variable is the log of the welfare weights assigned by Social Architects. Our explanatory variables include the log of the incomes of the Recipients, a set of dummies indicating Social Architects' beliefs about the needs and deservingness of high-income and low-income individuals, and the interaction terms of the log of the incomes of the Recipients with all the other variables. The base category takes a value of 1 if a Social Architect indicates that high-income individuals do not deserve and do not need their current income and low-income individuals do not deserve and do not need their current income.

The results are presented in Figure A12. We find that the coefficient estimate of *ln(recipient income) × HI deserve but do not need* is greater than 0. This implies that when Social Architects believe that high-income individuals deserve their current incomes (conditional on not needing them), they assign less progressive weights. We find a similar effect, qualitatively, for Social Architects who believe that high-income individuals deserve their incomes, conditional on needing them. This can be seen by comparing the coefficient estimates of *ln(recipient income) × HI do not deserve but need* and *ln(recipient income) × HI deserve and need*. From the coefficient estimate of *ln(recipient income) × HI do not deserve but need*, we find that when Social Architects believe that high-income individuals need their current incomes (conditional on not deserving them), they assign less progressive

weights. We find a similar effect, qualitatively, for Social Architects who believe that high-income individuals need their current incomes, conditional on deserving them. Social Architects' beliefs about the needs and deservingness of low-income individuals have a much smaller effect on their welfare weights.
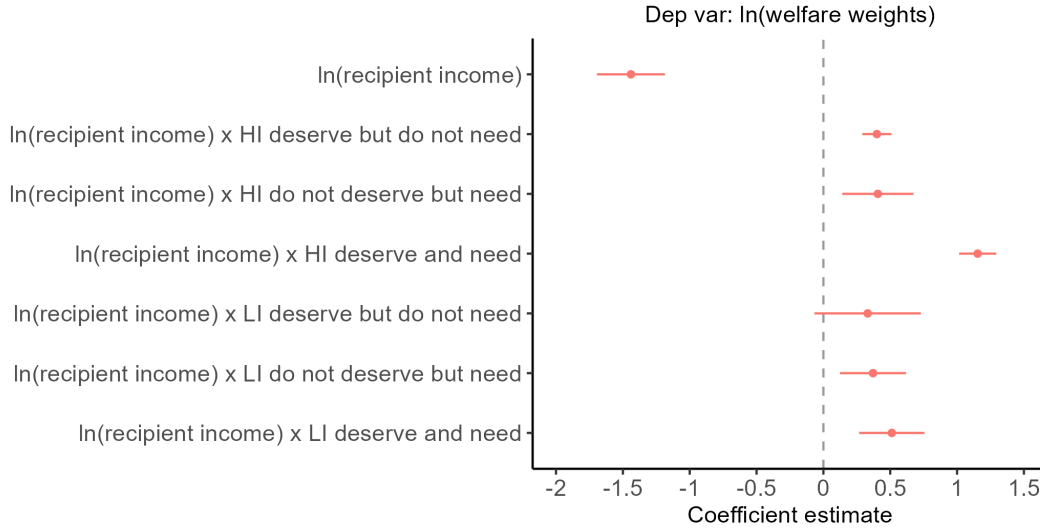
Figure A12: Welfare Weights and Non-Welfarist Ideals

Notes: The figure presents coefficient estimates. The dependent variable is the log of welfare weights assigned by Social Architects. The explanatory variables include the log of the incomes of the Recipients, a set of dummies indicating Social Architects' beliefs about the needs and deservingness of high-income and low-income individuals, and the interaction terms of the log of the incomes of the Recipients with all the other variables. Except for the main effect of the log of the incomes of the Recipients, we do not present the other main effects. *HI deserve but do not need* takes a value of 1 if a Social Architect indicates that high-income individuals deserve but do not need their current income. *LI deserve but do not need* takes a value of 1 if a Social Architect indicates that low-income individuals deserve but do not need their current income. The other variables are defined similarly. The base category takes a value of 1 if a Social Architect indicates that high-income individuals and low-income individuals do not deserve and do not need their current income. The regression is weighted using sampling weights. The regression uses data from Wave 2. Error bars are computed using HC3 standard errors.

Overall, we find that Social Architects' welfare weights capture welfarist and non-welfarist ideals.

D.2   Role of Social Architect's Own Income

To explore whether Social Architects assign higher welfare weights to Recipients with incomes similar to their own, we estimate the following regression.

$$g(R_j)_i = \beta_1 \textit{Income near R1}_{ij} + ... + \beta_7 \textit{Income near R7}_{ij} + \gamma_i + \epsilon_{ij}. \tag{A1}$$

The variable $g(R_j)_i$ is the weight assigned by Social Architect i to Recipient j. The

variable *Income near R1$_{ij}$* takes a value of 1 if Social Architect *i*'s income is near the income of Recipient 1. In particular, the variable takes a value of 1 if a Social Architect's income is less than or equal to $22,000. The other variables are similarly defined based on the following income brackets: ($22,000, $53,000], ($53,000, $85,000], ($85,000, $135,000], ($135,000, $210,000], ($210,000, $375,000], ($375,000, ∞). These income brackets are the same as the ones used in Treatments No Self-Interest and Self-Interest. We include Social Architect fixed effects γ_i to leverage the variation across decision screens. We exclude Treatment Self-Interest from the sample since this treatment incorporates self-interest motives by design. Here, we focus on understanding whether Social Architects are guided by self-interest motives in treatments designed to minimize the role of self-interest motives by design. We describe the role of self-interest motives in Treatment Self-Interest in Appendix Section D.3.

Figure A13 presents the coefficient estimates. Model s1 presents the specification described above. In Model s2, the variables *Income near R1$_{ij}$* through *Income near R7$_{ij}$* are indicator variables that take a value of 1 if Social Architect *i*'s income is within +- 20% of the income of Recipients 1 through 7, respectively.

We find that Social Architects with incomes near Recipient 1 assign a 10 percentage point higher weight to Recipient 1 relative to other Recipients. We find a similar effect for Social Architects with incomes near Recipients 2 and 4. Social Architects with incomes near Recipient 3 (earning $70,000) assign a lower weight to Recipient 3 relative to other Recipients. We do not find a statistically significant effect for Social Architects with incomes near Recipients 5, 6, and 7; we are likely underpowered to detect an effect for these Social Architects since our sample has very few high-income earning Social Architects.

D.3 Treatment Effects

Comparing Treatments Real and Hypothetical

We explore a possible reason why Social Architects might have more progressive welfare weights in Treatment Hypothetical relative to Treatment Real: Social Architects are less attentive in Treatment Hypothetical. Table A7 presents a series of regressions to investigate this hypothesis. The explanatory variable in each regression is a dummy variable that takes a value of 1 if a Social Architect is in Treatment Hypothetical and 0 if the Social Architect is in Treatment Real. In Column (1), the dependent variable is the time spent by the Social Architects on the survey. While Social Architects spend 1.6 fewer minutes on the survey in Treatment Hypothetical, this difference is not statistically significant. In Columns (2) and (3), the dependent variable takes a value of 1 if a Social Architect chooses the Constant Reform and the Variable Reform in each decision, respectively. We
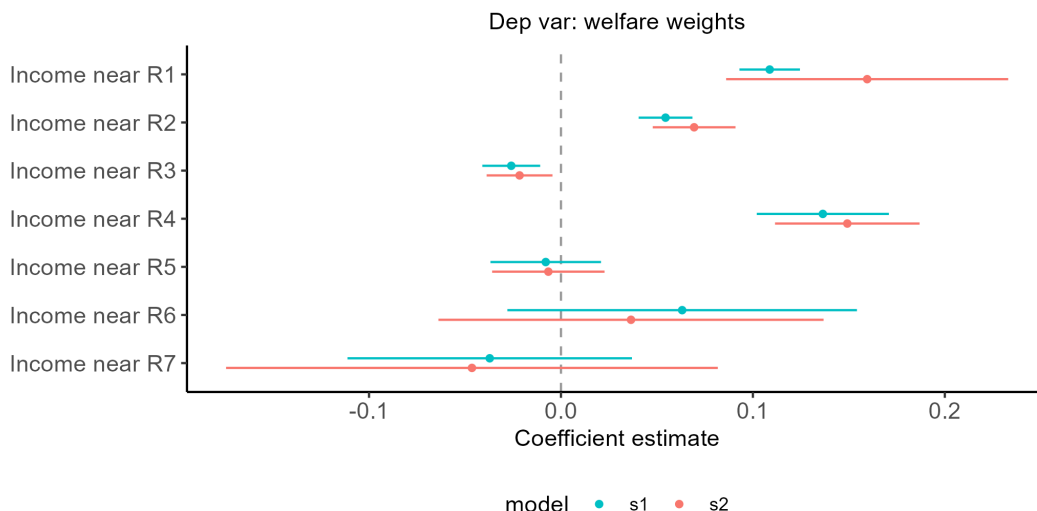
Figure A13: Weights Assigned to Recipients with Similar Incomes

Notes: The figure presents coefficient estimates. The dependent variable is the weights assigned by Social Architects. In Model s1, *Income near R1* takes a value of 1 if a Social Architect's income is less than or equal to $22,000. The other variables in Model s1 are similarly defined based on the following income brackets: ($22,000, $53,000], ($53,000, $85,000], ($85,000, $135,000], ($135,000, $210,000], ($210,000, $375,000], ($375,000, ∞). In model s2, *Income near R1* takes a value of 1 if a Social Architect's income is plus or minus 20% of the income of Recipient 1. The other variables in Model s2 are defined similarly. The regressions include Social Architect fixed effects and are weighted using sampling weights. The sample is restricted to all treatments, excluding Treatment Self-Interest. Error bars are computed using HC1 standard errors.

test whether Social Architects in Treatment Hypothetical choose one option throughout the survey to complete the survey quickly. We do not find evidence that the share of participants always choosing the Constant Reform or the Variable Reform is different in Treatment Hypothetical relative to Treatment Real. Overall, we do not find evidence that Social Architects in Treatment Hypothetical are less attentive in the survey.

Comparing Treatments Real and No-Self Interest

We explore the role of framing the Recipients' incomes by comparing Treatment No Self-Interest to Treatment Real. Treatments No Self-Interest is similar to Treatments Real, except that in the former, we provide Social Architects with the income brackets of the Recipients instead of the exact incomes of the Recipients. These income brackets span the income distribution.[48] We assume that Social Architects use the mean income of each income bracket as a reference when assigning welfare weights. These mean incomes are roughly similar to the incomes of the Recipients in Treatment Real. The results in Table 5 in the main text indicate that Social Architects are less progressive in Treatment No

---

[48]The income brackets are [0,$22,000], [$22,000, $53,000], [$53,000, $85,000], [$85,000, $135,000], [$135,000, $210,000], [$210,000, $375,000], [$375,000, ∞).

Table A7: Comparing Treatments Real and Hypothetical

Dependent Variables: Model:	Time spent (mins) (1)	Always CR (2)	Always VR (3)
Constant	16.758***	0.039***	0.119***
	(1.542)	(0.011)	(0.026)
Hypothetical	-1.679	0.018	-0.036
	(1.623)	(0.017)	(0.029)
Observations	997	997	997
Sampling Weights?	Yes	Yes	Yes

Notes: The table presents linear regressions. *Time spent (min)* is the time spent on the survey in minutes. *Always CR* (*Always VR*) is an indicator variable taking a value of 1 if a Social Architect chooses the Constant Reform (Variable Reform) in every decision. *Hypothetical* is an indicator variable taking a value of 1 if a Social Architect is in Treatment Hypothetical and a value of 0 if the Social Architect is in Treatment Real. The sample is restricted to Social Architects in Treatments Hypothetical and Real. The regressions weight each treatment using sampling weights. The Standard errors are robust to heteroskedasticity (HC3).
*p<0.1, **p<0.05, ***p<0.01

Self-Interest compared to Treatment Real (ν is 0.45 higher).

Treatment No Self-Interest does not control Social Architects' beliefs about Recipients' incomes. For example, Social Architects might believe that the Recipient in the "$375,000 and above" income bracket has an income of either $375,000 or $2,000,000. The estimated elasticity in this treatment can be sensitive to the assumptions about Social Architects' beliefs about Recipients' incomes.

Exploring Self-Interest Motives

In the main text, we found that participants in Treatment Self-interest have more progressive weights than participants in Treatment No Self-Interest, indicating that Social Architects' are guided by self-interest motives.In this section, we compare Social Architects' welfare weights in Treatment No Self-Interest and Self-Interest, separately by income groups. We pool the three groups with the highest incomes because we have a smaller share of participants in these groups. Figure A14 presents the results.

Considering the coefficient estimate of *ln(recipient income)* × *Self-Interest*, we find that for all income groups except the group earning above $135,000, the welfare weights are more progressive in Treatment Self-Interest relative to Treatment No Self-Interest. We are likely underpowered to detect effects for the group earning above $135,000. Overall, we find evidence that Social Architects across the income distribution are guided by self-
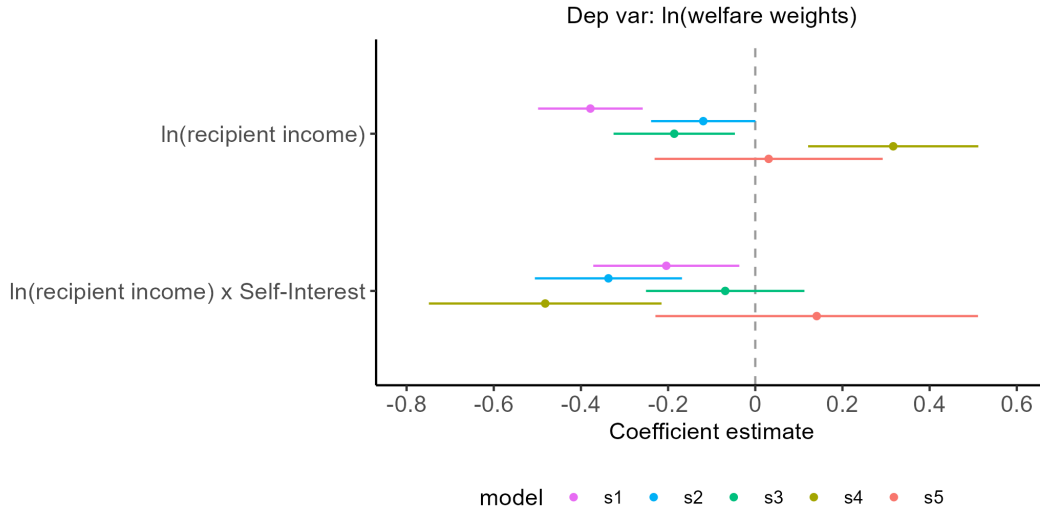
interest motives.

Figure A14: Social Architects' Weights and Self-Interest Motives

Notes: The figure presents coefficient estimates. The dependent variable is the log of the welfare weights assigned by Social Architects. The explanatory variables are the log of the incomes of the Recipients, *Self-Interest*, and the interaction term of the log of the incomes of the Recipients with *Self-Interest*. *Self-Interest* is a dummy variable taking a value of 1 if a Social Architect is in Treatment Self-Interest and 0 if a Social Architect is in Treatment No Self-Interest. The regressions use data from Treatments No Self-Interest and Treatment Self-Interest. Models s1 to s5 include Social Architects from the following income brackets [0, $22,000], [$22,000, $53,000], [$53,000, $85,000], [$85,000, $135,000], [$135,000, $210,000], [$210,000, $375,000], [$375,000, ∞), respectively. We do not present the main effects of *Self-Interest*. Error bars are computed using HC3 standard errors.

D.4   Welfare Weights and Support for Redistribution

We explore the empirical link between Social Architects' welfare weights and their support for government redistribution. In the main text, we explored the link using a question (*Redistribution*) administered in Wave 2. In this section, we explore the link using two validated measures of support of redistribution administered in Wave 1. The first question asks Social Architects if they think the government should do more to reduce income differences between the rich and poor. This question was used in the General Social Survey (GSS). The variable *Govt should do more* takes values from 1 to 7, with higher values indicating a greater desire for the government to do something to reduce inequality. The second question asks Social Architects if they want to increase or decrease the taxes on top-income earners. A modified version of this question was used by Cohn et al. (2019) and Kuziemko et al. (2015). The variable *Increase top-taxes* takes values from 1 to 7. A value below 4 indicates a desire to decrease top-taxes, while a value above 4 indicates a desire to increase top-taxes. A value of 4 indicates a desire to leave the top-taxes unchanged.

The order of the questions was counterbalanced across participants.

We benchmark the predictive power of Social Architects' welfare weights, measured by the elasticity of the weights, against their stated political affiliation. We estimate regressions in which the dependent variables are *Govt should do more* or *Increase top-taxes*. To assess the predictive power of a specification, we compute the root mean squared error (RMSE) of the out-of-sample predictions obtained from a specification.[49] The lower the RMSE of a specification, the higher the predictive power of that specification.

Table A8 presents the results. In rows one and two, we find that the elasticity of Social Architects' welfare weights is just as good a predictor of their support for redistribution as their stated political affiliation. In rows five and six, we find that the predictive power of welfare weights is slightly lower than that of political affiliation. However, the difference in the average prediction is small: the difference of 0.06 is 3% of the standard deviation (2.07) of the support for redistribution.

We find similar results in rows three, four, seven, and eight, which include treatment dummies and background characteristics as controls in the regressions.

D.5   Partisan Gap in Support for Redistribution

The results presented in the main text suggest that Republicans have less progressive support for government redistribution compared to Democrats and Independents. In this section, we explore the factors that explain the partisan gap in support for redistribution.

We explore the overall partisan gap in support for redistribution by regressing Social Architects' support for redistribution (variable *Redistribution*) on their political affiliation. The results are presented in Model s1 in Figure A15. The coefficient estimate of *Republican* in Model s1 is -0.69. This estimate reflects the overall partisan gap in support for redistribution.

Next, we explore the change in the coefficient estimate of *Republican* when we include other variables in the regression. Details about the other variables can be found in Appendix Section A. The results are presented in Model s2 in Figure A15a. We find that the coefficient estimate of *Republican* drops from -0.69 in Model s1 to -0.14 in Model s2 when we include other variables in the regression, indicating these variables explain 79.4% of the partisan gap in support for redistribution.

How much of the change in the coefficient estimate of *Republican* (between Model s1

---

[49]We divide the data into four sub-samples ($S(k), k \in 1, 2, 3, 4$) with $k = 4$. For each sub-sample, we train the specification of interest using the other three sub-samples ($S(-k)$). Next, we predict values for the sub-sample we left out and calculate the squared error, which is the difference between the actual and predicted values squared. To obtain the RMSE, we compute the square root of the average of the squared errors across all four sub-samples.

Table A8: Welfare Weights and Support for Redistribution

Row	Dependent variable	Explanatory variable	Controls?	RMSE
1	Increase top-taxes	Republican	No	1.57
2	Increase top-taxes	Elasticity	No	1.57
3	Increase top-taxes	Republican	Yes	1.55
4	Increase top-taxes	Elasticity	Yes	1.56
5	Govt should do more	Republican	No	2.00
6	Govt should do more	Elasticity	No	2.06
7	Govt should do more	Republican	Yes	1.97
8	Govt should do more	Elasticity	Yes	2.01
9	Increase top-taxes	Republican + Elasticity	No	2.00
10	Increase top-taxes	Republican + Elasticity	No	2.00

Notes: Each row of the table presents the root-mean-squared error (RMSE) of the predictions generated from a linear regression using a k-fold cross-validation procedure with $k = 4$. The dependent variable *Govt should do more* takes values from 1 through 7, with higher values indicating a greater desire for the government to do something to reduce inequality. The dependent variable *Increase top-taxes* takes values from 1 to 7, with higher values indicating a greater desire for the government to increase top-taxes. *Elasticity* is the elasticity of Social Architects' weights with respect to Recipients' incomes. *Republican* is a dummy variable taking a value of 1 for Republicans and a value of 0 for Democrats or Independents. The controls in the regression include a set of treatment dummies, *High Income* (= 1 if above median income), *Male* (=1 if male), *High Education* (=1 if above median education), and *High Age* (=1 if above median age), and *Policy Order* (dummy indicating the order of the two questions). The regressions use data from Wave 1.

and s2) is due to each of the other variables? Figure A15b presents the effect of each variable on the coefficient estimate of *Republican*. The effects are computed using the covariate decomposition procedure proposed by Gelbach (2016). A negative value for a variable indicates that its inclusion in the regression reduces the absolute value of the coefficient estimate *Republican* in Model s1. We find that about 8% of the partisan gap in support for redistribution is driven by welfare weights. Beliefs about higher taxes on high-income individuals hurting the economy and beliefs about the externalities from inequality explain 14% and 27% of the partisan gap, respectively.
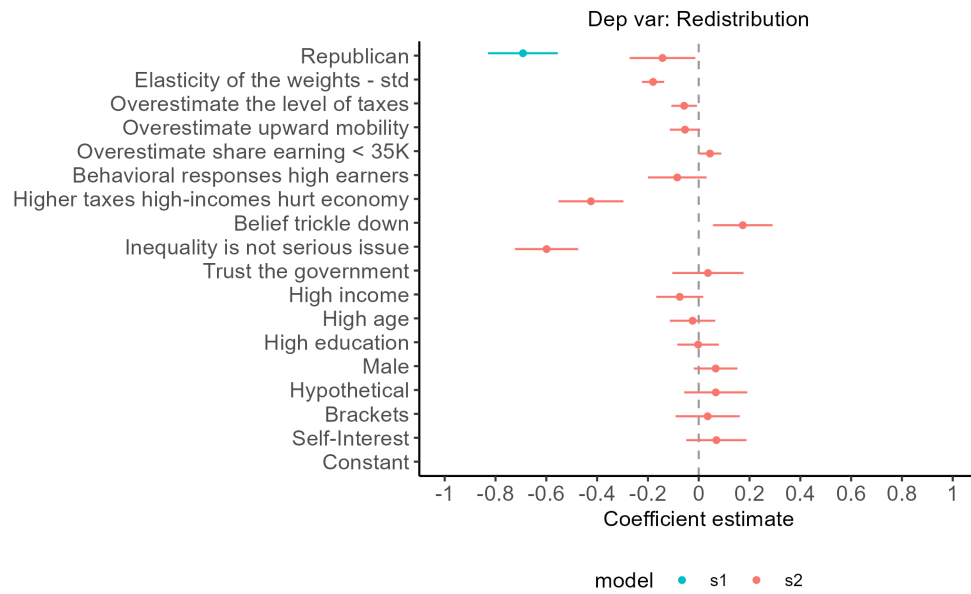
(a) Regression estimates
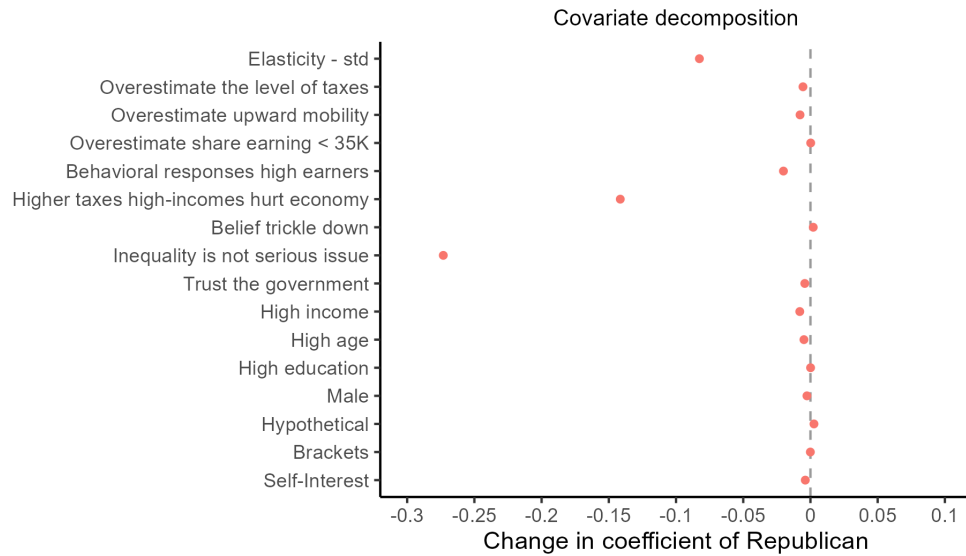
(b) Covariate Decomposition

Figure A15: Partisan Gap in Support for Redistribution

Notes: Panel (a) presents coefficient estimates. The dependent variable (*Redistribution*) takes values from -2 to +2, where positive (negative) values indicate that income should be further redistributed by taking from higher-income (lower/middle-income) individuals and giving to lower/middle-income (higher-income) individuals. A value of zero indicates that incomes should not be further redistributed. See the main text for an explanation of the explanatory variables. In these regressions, we reverse-code the variable *Inequality is a serious issue*. Models s1 and s2 are weighted using sampling weights. Error bars are computed using HC3 standard errors. Panel (b) presents the effect of each variable on the coefficient estimate of *Republican*. The effects are computed using the covariate decomposition procedure proposed by Gelbach (2016). The figures use data from Wave 2.

74

E Details on Analyses

E.1 Details on Plotting Recipients' incomes

We present the details of the construction of the figure that plots Recipients' incomes. Data on the income distribution is obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). Our measure of income is disposable incomes (*diinc*), which includes in-cash and in-kind transfers. For simplicity, every individual in the dataset is treated as a single filer, irrespective of their actual tax filing status.

Step 1: Exclude individuals with negative disposable incomes.

Step 2: Identify disposable income thresholds for percentiles (1 ..., 99).

Step 3: Assign the seven Recipients to the percentile matching their incomes.

Step 4: Plot the disposable income thresholds against the corresponding percentiles. Overlay the incomes of the Recipients.

E.2 Details on Plotting Welfare Weights

We present details on constructing the figures that plot welfare weights against the income distribution. Data on the income distribution is obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). Our measure of income is disposable incomes (*diinc*), which includes in-cash and in-kind transfers. For simplicity, every individual in the dataset is treated as a single filer, irrespective of their actual tax filing status.

Step 1: Exclude individuals with negative disposable incomes.

Step 2: Identify disposable income thresholds for percentiles (1 ..., 99).

Step 3: Compute the average disposable income in each percentile.

Step 4: Interpolate the welfare weights using the function y^v, where y is the average disposable income in each percentile, and v is a parameter. Different values of v lead to different welfare weights. Re-normalize the welfare weights so that they sum to 1.

Step 5: Plot the re-normalized welfare weights against the percentiles of the income distribution.

E.3 Estimating Likelihood of Being Registered Voters

We outline the methodology used to estimate the likelihood of Social Architects being registered voters. The data on voting behavior and demographics are obtained from the Cooperative Election Study (CCES) 2022 wave. The CCES is a 50,000+ person national stratified sample survey administered by YouGov.

Step 1: Estimate a logistic regression model to estimate the probability of an individual

being a registered voter. The predictors in the model include dummy variables indicating individuals' income, age, education, gender, region, and political affiliation. The set of predictors is the same as those indicated in Table 4, with the exception that we replace variables indicating individual income with variables indicating family income.

Step 2: Use the fitted logistic regression to obtain the predicted probability of an observation in our sample being a registered voter. Even though the model was trained using family income, the predictions are based on individual income because we only collected data on individual incomes in the survey.

E.4 Calibration of Optimal Income Taxes

We explore the implications of the different estimates of welfare weights for the optimal non-linear labor income taxes in the U.S. We use the optimal tax formula derived in Saez (2001). The formula is the solution to the planner's problem of maximizing social welfare, given the constraints on government revenue. It provides the optimal marginal tax rates for different incomes as a function of (i) the shape of the income distribution, (ii) welfare weights, and (iii) the elasticity of taxable income.

We use the actual income distribution in the U.S. to identify the shape of the income distribution. We use estimates of the elasticity of taxable incomes given in the literature. Finally, we use different estimates of welfare weights computed in our paper. Our calibrations follow the procedures of Mankiw et al. (2009) and Støstad and Cowell (2022).

Optimal Tax Formula

We assume that individuals are on a continuum of wage-earning abilities w with densities $f(w)$ and cumulative distribution function $F(w)$. Individuals' incomes are given by $z = wl$, where l is individuals' labor supply. Individuals' tax liabilities and consumption are denoted as $T(z)$ and $c = z - T(z)$. The planner observes income z and labor supply l, but does not know individuals' abilities w. We assume that every individual has additively separable preferences over consumption and labor given by

$$U(c,l) = u(c) + v(l) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{l^{1+\frac{1}{E_L}}}{(1+\frac{1}{E_L})}. \tag{A2}$$

These are constant relative risk aversion (CRRA) utilities in consumption with a coefficient of relative risk aversion given by γ. We assume a simple functional form for individuals' labor supply in which utilities are isoelastic in labor and depend on E_L, the elasticity of earnings with respect to the retention rate $1 - T'(z)$.

We denote the utility of an individual with wages w as $U(w)$. The planner has a

utilitarian social welfare function given by $W = \int U(\theta)f(\theta)d\theta$, where θ indexes wages in the integral. The planner's first-order condition is given by

$$\frac{T'(z(w))}{1 - T'(z(w))} = \left(\frac{1 + E_L^u(w)}{E_L^c(w)} \right) \frac{u'(c(w))}{wf(w)} \int_w^\infty \frac{1}{u'(c(\theta))} f(\theta)d\theta - (1 - F(w))\frac{1}{p}. \quad (A3)$$

where $T'(z(w))$ is the optimal marginal tax rate that applies to the income $z(w)$ at ability level w. The term $u'(c(w))$ is the marginal utility of consumption. It can be interpretted as welfare weights. Our assumption of CRRA utilities would result in $u'(c(w)) = c(w)^{1-\gamma}$. There is a mapping from the welfare weights obtained assuming CRRA utilities ($c^{1-\gamma}$) with a coefficient of relative risk aversion γ to the parametric form c^ν used in the paper to characterize welfare weights, given by $\gamma = -\nu$. Thus, in this approach, we implement different estimates of welfare wights using different values of γ.

The optimal tax rates are a function of $e^u(w)$ and $e^c(w)$, which are the average uncompensated and compensated elasticity of earnings with respect to the retention rate $1 - T'(z)$. The elasticity of earnings is assumed to be driven by the substitution effect (people work less due to increased taxes) and assumes no income effects (people work more due to increased taxes). Thus, we assume that $E_L^u(w) = E_L^c(w) = E_L$. The optimal marginal tax rate is decreasing in the elasticity of earnings to reduce the extent to which taxes distort people's labor supply.

In the above equation, p is the marginal value of public funds given by

$$p = \frac{1}{\int_0^\infty \frac{1}{u'(\theta)} f(\theta)d\theta} f(\theta)d\theta \quad (A4)$$

The term $u'(\theta)$ in Equation (A4) simplifies to the marginal utility of consumption $u'(c(\theta))$.

Estimating the Wage-Ability Distribution

The optimal taxes depends on individuals' incomes, which in turn depends on the taxes due to individuals' behavioral responses to taxes. Thus, we estimate the wage-ability distribution from the current income distribution and use this exogenous wage distribution when we calibrate the tax formula in Equation (A3). Data on the income distribution is obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). Each observation in the data corresponds to a tax unit.

Step 1: We use the NBER TAXSIM model to find the marginal tax rate for each tax unit.

The tax rates are calculated based on the available information about the tax units, which include the number of dependents, the age of the primary filer, and marital status. We then add a 5% state tax rate, a 2.9% tax rate for Medicare, and a 2.3% sales tax rate.

Step 2: Assuming that individuals have correctly optimized according to their utility function in Equation (A2), we back out the resulting wage-earning ability (w) of each tax unit.

Step 3: We create a smooth ability distribution using a Kernel density estimator with a bandwidth of $5000. The smooth distribution has 50,000 observations.

Step 4: We replace the top 0.5% of the distribution with a Pareto distribution. The Pareto parameter is the value of the local Pareto parameter $\alpha(z(w)) = \frac{z \cdot h(z)}{1 - H(z)}$ just before the top 0.5%.

Calibrating Optimal Income Taxes

We use an updating rule to find the fixed-point tax schedule. We assume an initial tax schedule. Given the tax schedule, we estimate individuals' labor supply. Individuals' labor supply is used to calculate their utilities. Given the utilities, we calculate the resulting optimal taxes at each wage level using Equation (A3). We iterate on this process until an optimum is found. We check if the second order condition holds at the optimum. This condition states that the pre-tax incomes are non-decreasing with wages.

Step 1: We start with an initial flat tax rate of 35%.

Step 2: We compute individual' labor choices based on this tax rate, assuming that they have correctly optimized their utilities in Equation (A2). Computing the derivative of Equation (A2) with respect to l and setting it to 0 yields $l = (w \cdot (1 - T'(z))^{E_L}$. We set $E_L = 0.25$, which is a mid-range estimate for the elasticity of taxable income (Saez et al., 2012).

Step 3: Based on the optimal labor choices computed in Step 2, we calculate the optimal income choices $z = wl$. We calculate individuals' utilities as based in Equation (A2). We use different values of γ in Equation (A2) to obtain different estimates of welfare weights. For example, we use $\nu = -0.1$ for the welfare weights implied by the income tax schedule in the U.S.

Step 5: We calculate the resulting optimal tax rate at each ability level based on Equation (A3).

Step 6: We repeat the previous steps until the tax rates converge to a fixed point.

F Instructions - Wave 1

Bold text, underlining, tables, etc., appear as on the original screens.

F.1 Treatment Loss x 70K

[Consent screen]

Introduction
Welcome to this research study. We appreciate your participation. We are a non-partisan group of researchers from University of Zurich and Erasmus University Rotterdam. This study contains real choices and questions regarding your demographic characteristics. No matter what your political views are, by completing this survey you are contributing to our knowledge as a society.

Time required
Approximately **10 minutes.** You will have a maximum of one hour to finish the survey after starting it.

Requirements
You must be a **U.S. resident** to participate in this study. You must also be above the age of 18. The survey contains attention checks. You must pass these check in order to proceed with the survey.

Confidentiality
All data obtained from you will be used for research purposes only. Data will be anonymized immediately after collection. Researchers will at no point have access to any information that could be used to personally identify you.

Voluntary participation
It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason.

Questions about the Survey
If you have questions about this study or your rights, please get in touch with us at Krishna.srinivasan@econ.uzh.ch

Consent
I have received the above information about the project and am willing to participate.

○ Yes; No

———————————————— page break ————————————————

[If a participant did not provide consent]

End of survey

You did not give your consent to continue with the study.

Thank you for your time.

You will be automatically redirected in 5 seconds.

———————————————— page break ————————————————

[Demographics screen]

What is your sex?

○ Male; Female

How old are you?

○ 18 years old - 34 years old; 35 years old - 44 years old; 45 years old - 54 years old; 55 years old - 64 years old; Above 65 years old

In which state do you currently reside?

○ Northeast (ME, NH, VT, MA, CT, RI, NY, PA, NJ); Midwest (OH, MI, IN, WI, IL, MN, IA, MO, ND, SD, NE, KS); South (DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX); Pacific (MT, WY, CO, NM, ID, UT, AZ, NV, WA, OR, CA, AK, HI); I do not reside in the US

What is the highest level of education you have completed?

○ Less than High School; High School/GED; Some College; Associate's Degree; Bachelor's degree; Master's degree; Doctoral or Profession Degree (PhD, ED.D, JD, DVM, DO, MD, DDS, or similar)

As of today, do you consider yourself a Republican, a Democrat, or an Independent?

○ Republican; Democrat; Independent

The next question is about your **total individual income in 2020 before taxes**. This figure should include income from all sources, including salaries, wages, pensions, Social Security, dividends, interest, and all other income. What was your total individual income (USD) in 2020?

- ○ $29,999 and below; $30,000 to $59,999; $60,000 to $99,999; $100,000 to $149,999; $150,00 and above

[Displayed if $29,999 and below is chosen]

You have reported that your total individual income in 2020 before taxes was $29,999 and below.

[Displayed if $30,000 to $59,999 is chosen]

You have reported that your total individual income in 2020 before taxes was between $30,000 and $59,999.

[Displayed if $60,000 to $99,999 is chosen]

You have reported that your total individual income in 2020 before taxes was between $60,000 and $99,999.

[Displayed if $100,000 to $149,999 is chosen]

You have reported that your total individual income in 2020 before taxes was between $100,000 and $149,999.

[Displayed if $150,000 and above is chosen]

You have reported that your total individual income in 2020 before taxes was above $150,000.

[Displayed in all cases]

Could you provide your best guess of what your **total individual income** was?

———————————————————— page break ————————————————————

[If quotas are full]

End of survey

Unfortunately, we already have the number of participants needed for this study.

Thank you for your time.

You will be automatically redirected in 5 seconds.

————————————————— page break —————————————————

[If a participant does not reside in the U.S]

End of survey

Unfortunately, you do not fulfil the requirements of this study since you do not reside in the U.S.

Thank you for your time.

You will be automatically redirected in 5 seconds.

————————————————— page break —————————————————

[Attention check screen]

In surveys like ours, some participants do not carefully read the questions. This means that there are a lot of random answers that can compromise the results of research studies. To show that you read our questions carefully, please choose both "Extremely interested" and "Not at all interested" below:

○ Extremely interested; Very interested; A little bit interested; Almost not interested; Not at all interested

————————————————— page break —————————————————

[If a participant failed the attention check]

End of survey

Sorry, you failed the attention check. You were supposed to select both "Extremely interested" and "Not at all interested."

You cannot continue with the study.

Thank you for your time.

You will be automatically redirected in 5 seconds.

———————————————————————— page break ————————————————————————

[Instructions screen]

Instructions

In this study, you will make several choices involving **seven real people**. These people will be selected at random from a survey panel and will not participate in the same survey as you. These people are above the age of 18 and are U.S. citizens. The incomes of the seven people are as follows:

Person	After-tax annual income
Person A	$8000
Person B	$35,000
Person C	$70,000
Person D	$100,000
Person E	$170,000
Person F	$250,000
Person G	$500,000

_____

Here is an example of a question that you will see in the survey:

	Person C	Person G
After-tax annual income	$70,000	$500,000

Question 2/4: Please choose your preferred alternative

Person C: +$750 Person G: -$1250	Person C: +$500 Person G: -$500

In this question, if you choose the option on the left, then $1250 will be taken away from Person G and $750 will be given to Person C. If you choose the option on the right, then $500 will be taken away from Person G and $500 will be given to Person C.

If you choose the option on the left, the final incomes of the two people **(including an initial $1500 bonus)** will be Person C: $72,250 and Person G: $500,250. If you choose the option on the right, the final incomes of the two people (including an initial $1500 bonus) will be Person C: $72,000 and Person G: $501,000.

83

You will face four questions like the one you saw above in each "decision screen." **Overall, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

There is a chance that you may be randomly selected in this study. If you are randomly selected, your choice on one randomly selected question on one randomly selected decision screen will be implemented. **This means that if you are randomly selected, one of your choices will have real consequences for two other people.** The final bonus of these two people will be transferred to them at the end of the study.

———

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving seven real people."

   ○ True; False

Please state True or False: "If you are randomly selected, one of your choices will have real consequences for two other people."

   ○ True; False

(You will be allowed to move to the next screen in 30 seconds)

———————————————————— page break ————————————————————

[If a participant fails the comprehension check]

End of survey

The correct answers were "True" and "True". You answered incorrectly.

You cannot continue with the study.

Thank you for your time.

You will be automatically redirected in 5 seconds.

———————————————————— page break ————————————————————

[Decision Screen 1 Question 1 (D1Q1): shown to all participants]

Please consider each question carefully because if you are selected, one of your choices
will have real consequences for two other persons.

	Person A	**Person G**
After-tax annual income	**$8,000**	**$500,000**

Question 1/4: Please choose your preferred alternative

Person A: +$1000 Person G: -$1000	Person A: +$500 Person G: -$500

———————————————— page break ————————————————

[All questions hereafter in Decision Screen 1 look like D1Q1]

[D1Q2.1: If (500, −500) chosen in D1Q1, choose between (1250, −750) and (500, −500)]

[D1Q2.2: If (1000, −1000) chosen in D1Q1, choose between (750, −1250) and (500, −500)]

———————————————— page break ————————————————

[D1Q3.1: If (500, −500) chosen in D1Q2.1, choose between (1375, −625) and (500, −500)]

[D1Q3.2: If (1250, −750) chosen in D1Q2.1, choose between (1125, −875) and (500, −500)]

[D1Q3.3: If (500, −500) chosen in D1Q2.2, choose between (875, −1125) and (500, −500)]

[D1Q3.4: If (750, −1250) chosen in D1Q2.2, choose between (625, −1375) and (500, −500)]

———————————————— page break ————————————————

[D1Q4.1: If (500, −500) chosen in D1Q3.1, choose between (1450, −550) and (500, −500)]

[D1Q4.2: If (1375, −625) chosen in D1Q3.1, choose between (1300, −700) and (500, −500)]

[D1Q4.3: If (500, −500) chosen in D1Q3.2, choose between (1200, −800) and (500, −500)]

[D1Q4.4: If (1125, −875) chosen in D1Q3.2, choose between (1050, −950) and (500, −500)]

[D1Q4.5: If (500, −500) chosen in D1Q3.3, choose between (950, −1050) and (500, −500)]

[D1Q4.6: If (875, −1125) chosen in D1Q3.3, choose between (800, −1200) and (500, −500)]

[D1Q4.7: If (500, −500) chosen in D1Q3.4, choose between (700, −1300) and (500, −500)]

[D1Q4.8: If (625, −1375) chosen in D1Q3.4, choose between (550, −1450) and (500, −500)]

———————————————————— page break ————————————————————

[Decision Screens 2-6 are identical to Decision Screen 1, with the exception that the incomes of the Recipients are different. The pair of Recipients they view is as follows:

B: $35,000 vs. C: $70,000 (Decision Screen 2)

C: $70,000 vs. D: $100,000 (Decision Screen 3)

C: $70,000 vs. E: $170,000 (Decision Screen 4)

C: $70,000 vs. F: $250,000 (Decision Screen 5)

C: $70,000 vs. G: $500,000 (Decision Screen 6)]

[For half the participants, the order of the Decision Screens is reversed. The pair of Recipients are as follows: C: $70,000 vs. G: $500,000 (Decision Screen 1), C: $70,000 vs. F: $250,000 (Decision Screen 2), C: $70,000 vs. E: $170,000 (Decision Screen 3), C: $70,000 vs. D: $100,000 (Decision Screen 4), B: $35,000 vs. C: $70,000 (Decision Screen 5), and A: $8,000 vs. C: $70,000 (Decision Screen 6).]

———————————————————— page break ————————————————————

[Policy views screen]

[The order of the two questions is counterbalanced across participants in each treatment.]

We have some final questions. It is important for us that you answer them carefully.

The top income tax category in 2020 includes those with an annual individual income of over $518,400. Do you think that income taxes levied on these people in the top income category should be increased, stay the same, or decreased?

    ○ 1 - Increased a lot

    ○ ...

    ○ 4 - Stay the same

- ...

- 7 - Decreased a lot

_____

Some people think that the government in Washington ought to reduce the income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing this income difference between the rich and the poor.

Here is a scale from 1 to 7. Think of a score of 1 as meaning that the government ought to reduce the income differences between rich and poor, and a score of 7 meaning that the government should not concern itself with reducing income differences. What score between 1 and 7 comes closest to the way you feel?

- 1 - Government should do something to reduce income differences between rich and poor

- ...

- 7 - Government should not concern itself with income differences

———————————————————— page break ————————————————————

End of survey

Thank you for your time!

You will be automatically redirected in 5 seconds.

F.2 Treatment Loss x 500K

[All screens, with the exceptions of those listed below, are identical to the screens in Treatment Loss x Moderate]

[Decision Screens 1 to 6 are identical to the corresponding Decision Screens in Treatment Loss x Moderate, with the exception that the incomes of the Recipients are different. The pair of Recipients they view is as follows:

A: $8,000 vs. G: $500,000 (Decision Screen 1)

B: $35,000 vs. G: $500,000 (Decision Screen 2)

C: $70,000 vs. G: $500,000 (Decision Screen 3)

D: $100,000 vs. G: $500,000 (Decision Screen 4)

E: $170,000 vs. G: $500,000 (Decision Screen 5)

F: $250,000 vs. G: $500,000 (Decision Screen 6)]

[For half the participants, the order of the Decision Screens is reversed]

F.3 Treatment Gain x 70K

[All screens, with the exceptions of those listed below, are identical to the screens in Treatment Loss x 70K]

[Instructions screen]

Instructions

In this study, you will make several choices involving **seven real people**. These people will be selected at random from a survey panel and will not participate in the same survey as you. These people are above the age of 18 and are U.S. citizens. The incomes of the seven people are as follows:

Person	After-tax annual income
Person A	$8000
Person B	$35,000
Person C	$70,000
Person D	$100,000
Person E	$170,000
Person F	$250,000
Person G	$500,000

_____

Here is an example of a question that you will see in the survey:

	Person C	Person G
After-tax annual income	$70,000	$500,000

Question 2/4: Please choose your preferred alternative

Person C: +$2250 Person G: +$250	Person C: +$2000 Person G: +$1000

In this question, if you choose the option on the left, then $250 will be given to Person G and $2250 will be given to Person C. If you choose the option on the right, then $1000 will be given to Person G and $2000 will be given to person C.

If you choose the option on the left, the final incomes of the two people will be Person C: $72,250 and Person G: $500,250. If you choose the option on the right, the final incomes of the two people will be Person C: $72,000 and Person G: $501,000.

You will face four questions like the one you saw above in each "decision screen." **Overall, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

There is a chance that you may be randomly selected in this study. If you are randomly selected, your choice on one randomly selected question on one randomly selected decision screen will be implemented. **This means that if you are randomly selected, one of your choices will have real consequences for two other people.** The final bonus of these two people will be transferred to them at the end of the study.

_____

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving seven real people."

- ○ True; False

Please state True or False: "If you are randomly selected, one of your choices will have real consequences for two other people."

- ○ True; False

(You will be allowed to move to the next screen in 30 seconds)

———————————————————————— page break ————————————————————————

[The incomes of the Recipients in the six decision screens are identical to the incomes of the Recipients in Treatment Loss x 70K.]

[Decision screen 1]

[D1Q1: Architect chooses between $(2500, 500)$ and $(2000, 1000)$]

———————————————————————— page break ————————————————————————

[D1Q2.1: If $(2000, 1000)$ chosen in D1Q1, choose between $(2750, 750)$ and $(2000, 1000)$]

[D1Q2.2: If $(2500, 500)$ chosen in D1Q1, choose between $(2250, 250)$ and $(2000, 1000)$]

[D1Q3.1: If $(2000, 1000)$ chosen in D1Q2.1, choose between $(2875, 875)$ and $(2000, 1000)$]

[D1Q3.2: If $(2750, 750)$ chosen in D1Q2.1, choose between $(2625, 625)$ and $(2000, 1000)$]

[D1Q3.3: If $(2000, 1000)$ chosen in D1Q2.2, choose between $(2375, 375)$ and $(2000, 1000)$]

[D1Q3.4: If $(2250, 250)$ chosen in D1Q2.2, choose between $(2125, 125)$ and $(2000, 1000)$]

[D1Q4.1: If $(2000, 1000)$ chosen in D1Q3.1, choose between $(2950, 950)$ and $(2000, 1000)$]

[D1Q4.2: If $(2875, 875)$ chosen in D1Q3.1, choose between $(2800, 800)$ and $(2000, 1000)$]

[D1Q4.3: If $(2000, 1000)$ chosen in D1Q3.2, choose between $(2700, 700)$ and $(2000, 1000)$]

[D1Q4.4: If $(2625, 625)$ chosen in D1Q3.2, choose between $(2550, 550)$ and $(2000, 1000)$]

[D1Q4.5: If $(2000, 1000)$ chosen in D1Q3.3, choose between $(2450, 450)$ and $(2000, 1000)$]

[D1Q4.6: If $(2375, 375)$ chosen in D1Q3.3, choose between $(2300, 300)$ and $(2000, 1000)$]

[D1Q4.7: If $(2000, 1000)$ chosen in D1Q3.4, choose between $(2200, 200)$ and $(2000, 1000)$]

[D1Q4.8: If $(2125, 125)$ chosen in D1Q3.4, choose between $(2050, 50)$ and $(2000, 1000)$]

[The questions in the other decision screens are identical to those in Decision Screen 1]

F.4   Treatment Gain x 500K

[All screens are identical to the screens in Treatment Gain x 70K, with the following exceptions: The incomes of the Recipients in the six decision screens are identical to the incomes of the Recipients in Treatment Loss x 500K.]

G   Instructions - Wave 2

Bold text, underlining, tables, etc., appear as on the original screen.

G.1   Treatment Real

This is an academic study conducted by the University of Zurich and Erasmus University Rotterdam.

- What you will do: You will make a number of decisions.

- Time required: Approximately 12 minutes.

- Requirements: In order to take part, you need to be a U.S. resident

———————————————————————— page break ————————————————————————

[Consent screen]

Introduction
Welcome to this research study. We appreciate your participation. We are a non-partisan group of researchers from University of Zurich and Erasmus University Rotterdam. This study contains real choices and questions regarding your demographic characteristics. No matter what your political views are, by completing this survey you are contributing to our knowledge as a society.

Time required
Approximately **12 minutes.**

Requirements
You must be a U.S. resident to participate in this study. You must also be above the age of 18. The survey contains attention checks. You must pass these check in order to proceed with the survey.

Confidentiality
All data obtained from you will be used for research purposes only. Data will be anonymized immediately after collection. Researchers will at no point have access to any information that could be used to personally identify you.

Voluntary participation

It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason.

Questions about the Survey

If you have questions about this study or your rights, please get in touch with us at Krishna.srinivasan@econ.uzh.ch

Consent

I have received the above information about the project and am willing to participate.

- ○ Yes; No

What is your prolific ID?

———————————————————— page break ————————————————————

[If a participant did not provide consent]

You did not give your consent to continue with the study.

Thank you for your time.

Please return your submission on Prolific by selecting the 'Stop without completing' button.

———————————————————— page break ————————————————————

[Demographics screen]

What is your sex?

- ○ Male; Female

How old are you?

- ○ 18 years old - 34 years old; 35 years old - 44 years old; 45 years old - 54 years old; 55 years old - 64 years old; 65 years old or above

In which state do you currently reside?

- ○ Alabama; ...; Wyoming; I do not reside in the U.S.

In which ZIP code do you live? (5 digits)

What is the highest level of education you have completed?

- Less than High School; High School/GED; Some College; Associate's Degree; Bachelor's degree; Master's degree; Doctoral or Profession Degree (PhD, ED.D, JD, DVM, DO, MD, DDS, or similar)

As of today, do you consider yourself a Republican, a Democrat, or an Independent?

- Republican; Democrat; Independent

The next question is about your **total individual income in 2021 before taxes**. This figure should include income from all sources, including salaries, wages, pensions, social security, dividends, interest, and all other income. What was your total individual income (USD) in 2021?

- $29,999 and below; $30,000 to $59,999; $60,000 to $99,999; $100,000 to $149,999; $150,00 and above

———————————————————— page break ————————————————————

[Displayed if $29,999 and below is chosen]

You have reported that your total individual income in 2021 before taxes was $29,999 and below.

[Displayed if $30,000 to $59,999 is chosen]

You have reported that your total individual income in 2021 before taxes was $30,000 to $59,999.

[Displayed if $60,000 to $99,999 is chosen]

You have reported that your total individual income in 2021 before taxes was $60,000 to $99,999.

[Displayed if $100,000 to $149,999 is chosen]

You have reported that your total individual income in 2021 before taxes was $100,000 to $149,999.

[Displayed if $150,000 and above is chosen]

You have reported that your total individual income in 2021 before taxes was $150,000 and above.

[Displayed in all cases]

Could you provide your best guess of what your **total individual income** was?

———————————————— page break ————————————————

[If a participant does not reside in the U.S]

End of survey

Unfortunately, you do not fulfil the requirements of this study since you do not reside in the U.S.

Thank you for your time.

Please return your submission on Prolific by selecting the 'Stop without completing' button.

———————————————— page break ————————————————

[Attention check screen]

In surveys like ours, some participants do not carefully read the questions. This means that there are a lot of random answers that can compromise the results of research studies. To show that you read our questions carefully, please choose both "Extremely interested" and "Not at all interested" below:

- ○ Extremely interested; Very interested; A little bit interested; Almost not interested; Not at all interested

———————————————— page break ————————————————

[Instructions screen]

Instructions

In this study, you will make several choices involving **seven real people**. These people will be selected at random from a survey panel and will not participate in the same survey as you. These people are above the age of 18 and are U.S. citizens. The incomes of the seven people **after all taxes paid and transfers received** are as follows:

Person	After-tax annual income
Person A	$8,000
Person B	$35,000
Person C	$70,000
Person D	$100,000
Person E	$170,000
Person F	$250,000
Person G	$500,000

Here is an example of a question that you will see in the survey:

	Person C	**Person G**
After-tax annual income	**$70,000**	**$500,000**

Question 2/4: Please choose your preferred alternative

Person C: +$750	Person C: +$500
Person G: -$1250	Person G: -$500

In this question, if you choose the option on the left, then $1250 will be taken away from Person G and $750 will be given to Person C. If you choose the option on the right, then $500 will be taken away from Person G and $500 will be given to Person C.

If you choose the option on the left, the final incomes of the two people **(including an initial $1500 bonus)** will be Person C: $72,250 and Person G: $500,250. If you choose the option on the right, the final incomes of the two people (including an initial $1500 bonus) will be Person C: $72,000 and Person G: $501,000.

You will face four questions like the one you saw above in each "decision screen." **Overall, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

One participant in this study will be randomly selected. If you are randomly selected, your choice on one randomly selected question on one randomly selected decision screen will be implemented. **This means that if you are randomly selected, one of your choices will have real consequences for two other people.** The final bonus of these two people will be transferred to them at the end of the study.

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving seven real people."

○ True; False

Please state True or False: "If you are randomly selected, one of your choices will have real consequences for two other people."

○ True; False

(You will be allowed to move to the next screen in 30 seconds)

[The timer updates dynamically. When the time elapses, the text disappears.]

———————————————— page break ————————————————

[If a participant fails at least two out of three checks (one attention check and two comprehension checks)]

End of survey

Sorry, you answered at least two out of three comprehension/attention checks incorrectly.

You cannot continue with the study.

Thank you for your time.

Please return your submission on Prolific by selecting the 'Stop without completing' button.

[If a participant fails only one out of three checks (one attention check and two comprehension checks)]

End of survey

Thank you for your time.

We will pay you your £2 participation fee in the following days.

Please click the following link to finish the survey.

———————————————— page break ————————————————

[D1Q1: shown to all participants]

Please consider each question carefully because if you are selected, one of your choices will have real consequences for two other persons.

	Person A	**Person C**
After-tax annual income	**$8,000**	**$70,000**

Question 1/4: Please choose your preferred alternative:

Person A: +$1000	Person A: +$500
Person C: -$1000	Person C: -$500
○	○

————————————————— page break —————————————————

[All questions hereafter in Decision Screen 1 look like D1Q1]

[D1Q2.1: If $(500, -500)$ chosen in D1Q1, choose between $(1250, -750)$ and $(500, -500)$]

[D1Q2.2: If $(1000, -1000)$ chosen in D1Q1, choose between $(750, -1250)$ and $(500, -500)$]

————————————————— page break —————————————————

[D1Q3.1: If $(500, -500)$ chosen in D1Q2.1, choose between $(1375, -625)$ and $(500, -500)$]

[D1Q3.2: If $(1250, -750)$ chosen in D1Q2.1, choose between $(1125, -875)$ and $(500, -500)$]

[D1Q3.3: If $(500, -500)$ chosen in D1Q2.2, choose between $(875, -1125)$ and $(500, -500)$]

[D1Q3.4: If $(750, -1250)$ chosen in D1Q2.2, choose between $(625, -1375)$ and $(500, -500)$]

————————————————— page break —————————————————

[D1Q4.1: If $(500, -500)$ chosen in D1Q3.1, choose between $(1450, -550)$ and $(500, -500)$]

[D1Q4.2: If $(1375, -625)$ chosen in D1Q3.1, choose between $(1300, -700)$ and $(500, -500)$]

[D1Q4.3: If $(500, -500)$ chosen in D1Q3.2, choose between $(1200, -800)$ and $(500, -500)$]

[D1Q4.4: If (1125, −875) chosen in D1Q3.2, choose between (1050, −950) and (500, −500)]

[D1Q4.5: If (500, −500) chosen in D1Q3.3, choose between (950, −1050) and (500, −500)]

[D1Q4.6: If (875, −1125) chosen in D1Q3.3, choose between (800, −1200) and (500, −500)]

[D1Q4.7: If (500, −500) chosen in D1Q3.4, choose between (700, −1300) and (500, −500)]

[D1Q4.8: If (625, −1375) chosen in D1Q3.4, choose between (550, −1450) and (500, −500)]

————————————————— page break —————————————————

[Decision Screens 2-6 are identical to Decision Screen 1, with the exception that the incomes of the Recipients are different. The pair of Recipients they view is as follows:

B: $35,000 vs. C: $70,000 (Decision Screen 2)

C: $70,000 vs. D: $100,000 (Decision Screen 3)

C: $70,000 vs. E: $170,000 (Decision Screen 4)

C: $70,000 vs. F: $250,000 (Decision Screen 5)

C: $70,000 vs. G: $500,000 (Decision Screen 6)]

[For half the participants, the order of the Decision Screens is reversed.]

————————————————— page break —————————————————

How confident are you that the choices you made in the previous screens reflect what you really think?

Please provide your answer on a scale of 1 to 5. A 1 indicates "Not confident all," and a 5 indicates "Completely confident."

  ○ 5: Completely confident; 4:; 3:; 2:; 1: Not confident at all

————————————————— page break —————————————————

In the following screens, we would like to ask you some general questions about your views on society. Your opinion and thoughts are important to us.

Consider the current incomes of individuals in society obtained after all taxes are paid

and transfers received.

Which of the following statements comes closest to how you feel?

High-income individuals ...

- do <u>not</u> deserve their current income and do <u>not</u> need their current income

- deserve their current income but do <u>not</u> need their current income

- do <u>not</u> deserve their current income but need their current income

- deserve their current income and need their current income

_____

Which of the following statements comes closest to how you feel?

Low-income individuals ...

- do <u>not</u> deserve their current income and do <u>not</u> need their current income

- deserve their current income but do <u>not</u> need their current income

- do <u>not</u> deserve their current income but need their current income

- deserve their current income and need their current income

———————————————————————— page break ————————————————————————————

Consider the current incomes of individuals in society obtained after all taxes are paid and transfers received.

Do you think that, given the current incomes of individuals in society, incomes should be further redistributed or should not be further redistributed?

Please provide your answer on a scale from -2 to +2 where a +2 means that income should be further redistributed by taking from the higher-income individuals and giving to the lower/middle-income individuals while a -2 means that income should be further redistributed by taking from the lower/middle-income individuals and giving to the higher-income individuals.

- -2: Incomes should be further redistributed by taking from the lower/middle-income individuals and giving to the higher-income individuals

- ○ -1:

- ○ +0: Incomes should **not** be further redistributed

- ○ +1:

- ○ +2: Incomes should be further redistributed by taking from the higher-income individuals and giving to the lower/middle-income individuals

———————————————————— page break ————————————————————

The next set of questions is about the income tax system in the United States. These are questions for which there are right or wrong answers.

In order for your answers to be most helpful to us, it is really important that you answer these questions as accurately as you can. Although you may find some questions difficult, it is very important for our research that you try your best. Thank you very much!

———

Out of 100 households in the U.S., how many are in the top federal personal income tax bracket?

[slider 0-100]

———

What share of their total income do people in the top federal personal income tax bracket pay in taxes?

[slider 0-100]

———

Out of 100 U.S. households, how many pay no federal income taxes?

[slider 0-100]

———

Imagine a middle class household that is right at the middle of the income distribution, such that half of all households in the U.S. earn more than this household and half earn less. What share of their income do you think such a household pays in federal income taxes?

[slider 0-100]

Out of every 100 individuals in the U.S., how many earn an income (after all taxes paid and transfers received) below $35,000?

[slider 0-100]

We would now like to ask you what you think about the life opportunities of children from very poor families.

For the following question, we focus on 500 families that represent the U.S. population. We divide them into five groups on the basis of their income, with each group containing 100 families. These groups are:

- The poorest 100 families

- The second poorest 100 families

- The middle 100 families

- The second richest 100 families

- The richest 100 families

How many out of 100 children coming from the poorest 100 families will grow up to be among the richest 100 families?

——————————————————— page break ———————————————————

[Tax preferences screen]

We would like to ask you what you think the distribution of after-tax income in the U.S. should be.

There are **7 tax groups** (tax brackets) in the U.S. Group 1 includes households with the lowest incomes and Group 7 includes households with the highest incomes. Groups 2 through 6 include households with incomes in the middle.

Column 2 of the table below lists the **CURRENT** average annual after-tax income of all households in each group. The after-tax income is obtained by subtracting all federal income taxes (e.g., ordinary income taxes, alternative minimum taxes) from the pre-tax income and adding all federal transfers (e.g., tax credits) to the pre-tax income.

In Column 3 of the table below, we list the average federal income tax rate of each group. This rate was determined based on the ordinary income taxes that households paid. As

an example, if a household with a pre-tax income of $80,000 has an average tax rate of 15%, they would pay 80000*0.15 = $12,000 in taxes.

We would like you to indicate what you think the average tax rate for each tax group in the U.S. should be. This can be done as follows. **You can increase or decrease the average tax rates of the first six groups. The average tax rate of group 7 adjusts automatically so that all seven groups together pay as much taxes as they currently do.**

Column 4 of the table below and the figure below indicate your **DESIRED** average annual after-tax incomes. The numbers in the table as well as the figure update automatically as you change the average tax rates.

Your choices will sometimes be limited for a variety of reasons. For example, you cannot set the tax rate for a group such that their average after-tax income becomes lower than the average after-tax income of the group below them or higher than the average after-tax income of the group above them.

Note also that there may be rounding-off errors in various calculations.

You can go back to the initial situation by refreshing the page.

DESIRED Income Distribution

Highcharts.com
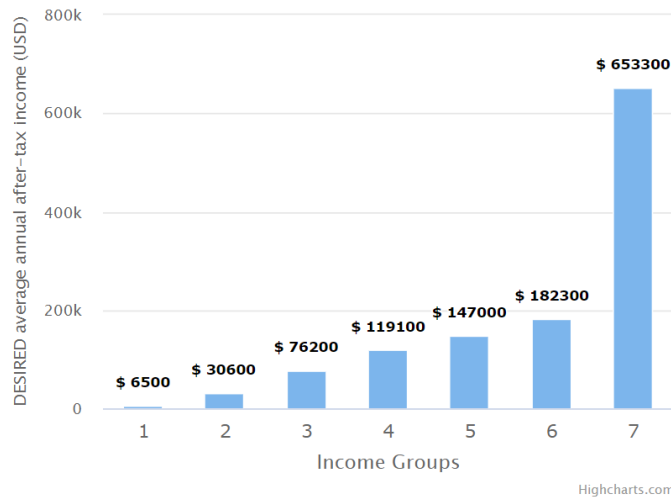
Income group	Annual after-tax income (CURRENT)	Average tax rate	Annual after-tax income (DESIRED)
1	$6,500	9%	$6,500
2	$30,600	11%	$30,600
3	$76,200	15%	$76,200
4	$119,100	19%	$119,100
5	$147,000	21%	$147,000
6	$182,300	25%	$182,300
7	$653,300	31%	$653,300

Please answer the following last set of questions.

Which has more to do with why a person is rich?

- ○ Because she or he worked harder than others; Because she or he had more advantages than others

———————————————————————————————————

If the federal personal income tax rate were to increase for the richest people in the economy, to what extent would it encourage them to work less?

- ○ A great deal; A lot; A moderate amount; A little; None at all

———————————————————————————————————

Do you think that increasing income taxes on high-income households would hurt economic activity, not have an effect on economic activity, or help economic activity in the U.S.?

- ○ Hurt economic activity in the U.S.; Not have an effect on economic activity in the U.S.; Help economic activity in the U.S.

———————————————————————————————————

Typically, when the top federal income tax rate on high earners is cut, do you think that the lower class and working class mostly win or mostly lose from this change?

- ○ Mostly lose; Neither lose nor win; Mostly win

———————————————————————————————————

Some people think that income inequality in society can affect the level of crime, trust, corruption, and social unrest in society.

How big of an issue do you think income inequality is in America?

- ○ Not an issue at all; A small issue; An issue; A serious issue; A very serious issue

———————————————————————————————————

How much of the time do you think you can trust the federal government to do what is right?

○ Always; Most of the time; Only some times; Never

———————————————————— page break ————————————————————

End of survey

Thank you for your time!

We will pay you your £2 participation fee in the following days.

Please click the following link to finish the survey.

G.2 Treatment Hypothetical

[All screens, with the exceptions of those listed below, are identical to the screens in Treatment Real]

[Instructions screen]

Instructions

In this study, you will make several choices involving **seven hypothetical people**. These people are not real but you should imagine them as above the age of 18 and U.S. citizens. The incomes of the seven people **after all taxes paid and transfers received** are as follows:

Person	After-tax annual income
Person A	$8,000
Person B	$35,000
Person C	$70,000
Person D	$100,000
Person E	$170,000
Person F	$250,000
Person G	$500,000

Here is an example of a question that you will see in the survey:

	Person C	Person G
After-tax annual income	**$70,000**	**$500,000**

Question 2/4: Please choose your preferred alternative

Person C: +$750	Person C: +$500
Person G: -$1250	Person G: -$500

In this question, if you choose the option on the left, then $1250 will be taken away from Person G and $750 will be given to Person C. If you choose the option on the right, then $500 will be taken away from Person G and $500 will be given to Person C.

If you choose the option on the left, the final incomes of the two people **(including an initial $1500 bonus)** will be Person C: $72,250 and Person G: $500,250. If you choose the option on the right, the final incomes of the two people (including an initial $1500 bonus) will be Person C: $72,000 and Person G: $501,000.

You will face four questions like the one you saw above in each "decision screen." **Over-all, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

The choices you make in the survey will not have real consequences.

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving seven hypothetical people."

○ True; False

Please state True or False: "Your choices will **not** have real consequences."

○ True; False

(You will be allowed to move to the next screen in 30 seconds)

[The timer updates dynamically. When the time elapses, the text disappears.]

———————————————————— page break ————————————————————

[D1Q1: shown to all participants]

Decision Screen 1/6

Please consider each question carefully.

	Person A	Person C
After-tax annual income	$8,000	$70,000

Question 1/4: Please choose your preferred alternative:

Person A: +$1000	Person A: +$500
Person C: -$1000	Person C: -$500
○	○

[All decision screens and questions and identical to those in Treatment Real. Only the first sentence differs between the two treatments]

G.3   Treatment No Self-Interest

[All screens, with the exceptions of those listed below, are identical to the screens in Treatment Real]

[In the Demographics screen, all questions, with the exception of the question on own income, is the same as in Treatment Real]

The next question is about your **total individual income in 2021 before taxes**. This figure should include income from all sources, including salaries, wages, pensions, Social Security, dividends, interest, and all other income. What was your total individual income (USD) in 2021?

   ○ $22,000 and below; $22,000 to $53,000; $53,000 to $85,000; $85,000 to $135,000; $135,000 to $210,000; $210,000 to $375,000; $375,000 and above

———————————————————— page break ————————————————————

[Displayed if $22,000 and below is chosen]

You have reported that your total individual income in 2021 before taxes was $22,000 and below.

[Displayed if $22,000 to $53,000 is chosen]

You have reported that your total individual income in 2021 before taxes was $22,000 to $53,000.

[Displayed if $53,000 to $85,000 is chosen]

You have reported that your total individual income in 2021 before taxes was $53,000 to $85,000.

[Displayed if $85,000 to $135,000 is chosen]

You have reported that your total individual income in 2021 before taxes was $85,000 to $135,000.

[Displayed if $135,000 to $210,000 is chosen]

You have reported that your total individual income in 2021 before taxes was $135,000 to $210,000.

[Displayed if $210,000 to $375,000 is chosen]

You have reported that your total individual income in 2021 before taxes was $210,000 to $375,000.

[Displayed if $375,000 and above is chosen]

You have reported that your total individual income in 2021 before taxes was $375,000 and above.

[Displayed in all cases]

Could you provide your best guess of what your **total individual income** was?

————————————————— page break —————————————————

[Instructions screen]

Instructions

In this study, you will make several choices involving **seven real people**. These people will be selected at random from a survey panel and will not participate in the same survey as you. These people are above the age of 18 and are U.S. citizens. The incomes of the seven people **after all taxes paid and transfers received** put them in the following income brackets:

Person	After-tax annual income
Person A	$22,000 and below
Person B	$22,000 to $53,000
Person C	$53,000 to $85,000
Person D	$85,000 to $135,000
Person E	$135,000 to $210,000
Person F	$210,000 to $375,000
Person G	$375,000 and above

Here is an example of a question that you will see in the survey:

	Person C	Person G
After-tax annual income	$53,000 to $85,000	$375,000 and above

110

Question 2/4: Please choose your preferred alternative

Person C: +$750	Person C: +$500
Person G: -$1250	Person G: -$500

In this question, if you choose the option on the left, then $1250 will be taken away from Person G and $750 will be given to Person C. If you choose the option on the right, then $500 will be taken away from Person G and $500 will be given to Person C.

If you choose the option on the left, the final income brackets of the two people **(including an initial $1500 bonus)** will be Person C: $55,250 to $87,250 and Person G: $375,250 and above. If you choose the option on the right, the final incomes of the two people (including an initial $1500 bonus) will be Person C: $55,000 to $87,000 and Person G: $376,000 and above.

You will face four questions like the one you saw above in each "decision screen." **Overall, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

One participant in this study will be randomly selected. If you are randomly selected, your choice on one randomly selected question on one randomly selected decision screen will be implemented. **This means that if you are randomly selected, one of your choices will have real consequences for two other people.** The final bonus of these two people will be transferred to them at the end of the study.

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving seven real people."

○ True; False

Please state True or False: "If you are randomly selected, one of your choices will have real consequences for two other people."

○ True; False

(You will be allowed to move to the next screen in 30 seconds)

[The timer updates dynamically. When the time elapses, the text disappears.]

———————————————————————— page break ————————————————————————

[D1Q1: shown to all participants]

Decision Screen 1/6

Please consider each question carefully because if you are selected, one of your choices will have real consequences for two other persons.

	Person A	Person C
After-tax annual income	$22,000 and below	$53,000 to $85,000

Question 1/4: Please choose your preferred alternative:

Person A: +$1000 Person C: -$1000 ○	Person A: +$500 Person C: -$500 ○

[All questions are identical to those in Treatment Real. Decision Screens 1 to 6 are identical to the corresponding Decision Screens in Treatment Real, with the exception that the incomes of the Recipients are different. The pair of Recipients they view is as follows:

Decision Screen 2 (B: $22,000 to $53,000 and C: $53,000 to $85,000)

Decision Screen 3 (C: $53,000 to $85,000 and D: $85,000 to $135,000)

Decision Screen 4 (C: $53,000 to $85,000 and E: $135,000 to $210,000)

Decision Screen 5 (C: $53,000 to $85,000 and F: $210,000 to $375,000)

Decision Screen 6 (C: $53,000 to $85,000 and G: $375,000 and above)]

[For half the participants, the order of the Decision Screens is reversed]

G.4 Treatment Self-Interest

[All screens, with the exceptions of those listed below, are identical to the screens in Treatment No Self-Interest]

[Instructions screen]

Instructions

In this study, you will make several choices involving **six real people** and you. These six people will be selected at random from a survey panel and will not participate in the same survey as you. These people are above the age of 18 and are U.S. citizens. The incomes of the six people **after all taxes paid and transfers received** put them in the following income brackets:

Note that in this study, you are Person [A/B/C/D/E/F/G] earning [income].

Person	After-tax annual income
Person A	$22,000 and below
Person B	$22,000 to $53,000
Person C	$53,000 to $85,000
Person D	$85,000 to $135,000
Person E	$135,000 to $210,000
Person F	$210,000 to $375,000
Person G	$375,000 and above

Here is an example of a question that you will see in the survey:

	Person C	**Person G**
After-tax annual income	**$53,000 to $85,000**	**$375,000 and above**

Question 2/4: Please choose your preferred alternative

Person C: +$750	Person C: +$500
Person G: -$1250	Person G: -$500

In this question, if you choose the option on the left, then $1250 will be taken away from Person G and $750 will be given to Person C. If you choose the option on the right, then $500 will be taken away from Person G and $500 will be given to Person C.

If you choose the option on the left, the final income brackets of the two people **(including an initial \$1500 bonus)** will be Person C: \$55,250 to \$87,250 and Person G: \$375,250 and above. If you choose the option on the right, the final incomes of the two people (including an initial \$1500 bonus) will be Person C: \$55,000 to \$87,000 and Person G: \$376,000 and above.

You will face four questions like the one you saw above in each "decision screen." **Overall, you will face six decision screens with four questions in each.** In each question, you will see a different amount in the option on the left. In each decision screen, you will see a different pair of people.

Remember that in this study, you are Person [A/B/C/D/E/F/G] earning [income].

One participant in this study will be randomly selected. If you are randomly selected, your choice on one randomly selected question on one randomly selected decision screen will be implemented. **This means that if you are randomly selected, one of your choices will have real consequences. If the selected question involves a payment to you, then we will pay out the bonus to you and to the other person. If the selected question involves a payment to two other persons, then we will pay out the bonus to these two other persons.** The final bonus will be transferred at the end of the study. If you are among the winners, we will contact you in a few months and pay out your bonus via prolific.

Please answer the following questions to demonstrate that you have understood the instructions. You can read the instructions above again if you feel the need to.

Please state True or False: "In this study, you will make several choices involving six real people and you."

○ True; False

Please state True or False: "If you are randomly selected, one of your choices will have real consequences for two other people or for you and one other person."

○ True; False

(You will be allowed to move to the next screen in 30 seconds)

[The timer updates dynamically. When the time elapses, the text disappears.]

———————————————— page break ————————————————

[D1Q1: shown to all participants]

Decision Screen 1/6

Please consider each question carefully because if you are selected, one of your choices will have real consequences.

	Person A	Person C
After-tax annual income	$22,000 and below	$53,000 to $85,000

Question 1/4: Please choose your preferred alternative:

Person A: +$1000 Person C: -$1000 ○	Person A: +$500 Person C: -$500 ○

[All questions and decision screens are identical to those in Treatment No Self-Interest with the exception that in the relevant decision screens, we replace "Person [A/B/C/D/E/F/G]" with "You." Furthermore, the first sentence in all decision screens is different.]

References

Alesina, A., Stantcheva, S. and Teso, E. (2018), 'Intergenerational mobility and preferences for redistribution', *American Economic Review* **108**(2), 521–554.

Cohn, A., Jessen, L. J., Klasnja, M. and Smeets, P. (2019), Why do the rich oppose redistribution? An experiment with america's top 5%. Working paper.

Gelbach, J. B. (2016), 'When do covariates matter? and which ones, and how much?', *Journal of Labor Economics* **34**(2), 509–543.

Hendren, N. (2020), 'Measuring economic efficiency using inverse-optimum weights', *Journal of Public Economics* **187**, 104198.

Hendren, N. and Sprung-Keyser, B. (2020), 'A unified welfare analysis of government policies', *The Quarterly Journal of Economics* **135**(3), 1209–1318.

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.

Lobeck, M. and Støstad Nyborg, M. (2022), 'The consequences of inequality: Beliefs and redistributive preferences', *Working paper* .

Mankiw, N. G., Weinzierl, M. and Yagan, D. (2009), 'Optimal taxation in theory and practice', *Journal of Economic Perspectives* **23**(4), 147–174.

Piketty, T., Saez, E. and Zucman, G. (2018), 'Distributional national accounts: methods and estimates for the united states', *The Quarterly Journal of Economics* **133**(2), 553–609.

Saez, E. (2001), 'Using elasticities to derive optimal income tax rates', *The review of economic studies* **68**(1), 205–229.

Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The elasticity of taxable income with respect to marginal tax rates: A critical review', *Journal of economic literature* **50**(1), 3–50.

Stantcheva, S. (2020), 'Understanding economic policies: What do people know and learn', *Unpublished Manuscript, Harvard University* pp. 2309–2369.

Stantcheva, S. (2021), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 2309–2369.

Støstad, M. N. and Cowell, F. (2022), 'Inequality as an externality: Consequences for tax design', *Available at SSRN 4185685* .