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

Francesco Capozza†
Krishna Srinivasan‡

August 8, 2023

[Most recent version]

Abstract

Evaluating the desirability of a reform involves weighing the gains of the winners against the losses of the losers using welfare weights. Welfare weights measure the value society places on providing an additional dollar of consumption to individuals. Which welfare weights should be used to evaluate policies? We take a positive approach to answer this question by identifying the welfare weights assigned by the general population of the U.S. using a real-stakes experiment. These weights are general enough to capture various ideals, such as equality of opportunity or poverty alleviation, and can be directly used to evaluate Social Architects' assessment of the desirability of reforms. We find that the average welfare weights of the general population are progressive, indicating a desire for 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 tax and transfer policies in the U.S. but less progressive than the weights frequently used in the optimal policy literature.

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

*This work has been supported by Grant 2104-31756 from the Russell Sage Foundation and by the University of Zurich's Research Priority Program "Equality of Opportunity." Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of the Russel Sage Foundation. We thank Stefanie Stantcheva, Nathaniel Hendren, Luca Henkel, Dmitry Taubinsky, Itai Sher, and Nina Serdarevic, for their helpful comments. We also thank seminar participants at Erasmus University Rotterdam, University of Zurich, U.C. Berkeley, National University of Singapore, and Norwegian School of Economics for their helpful comments. We are extremely grateful to Sandro Ambühl for his valuable comments and insights. This study was pre-registered: https://doi.org/10.1257/rct.8372-3.2

†Erasmus University Rotterdam. Email: capozza@ese.eur.nl
‡University of Zurich. Email: krishna.srinivasan@econ.uzh.ch.

1 Introduction

Policy reforms in domains such as income taxation, cash transfers, and in-kind transfers typically have winners and losers. Evaluating the desirability of a reform involves weighing the gains of the winners against the losses of the losers using welfare weights. Welfare weights measure the value society places on providing an additional dollar of consumption to individuals. Which welfare weights should be used to evaluate policies? The existing literature on optimal policies frequently assumes a functional form for the welfare weight. In our project, we identify the welfare weights assigned by the general population of the U.S. This is a valuable positive exercise since we can test if the welfare weights of the general population are similar to the welfare weights commonly used in the literature and the welfare weights implied by the existing tax and transfer system in the U.S. This is also a useful normative exercise since the elicited weights can be used to evaluate reforms to ensure their social acceptability.

To elicit the welfare weights of the general population of the U.S., we conducted two real-stakes online experiments ($N \approx 4000$) with samples of the general population in the U.S. An experimental approach allows us to elicit the welfare weights of the general population unconfounded by their views about the government and taxation. In the experiments, participants in the role of "Social Architects" face pairs of real "Recipients" whose disposable incomes span the income distribution of the U.S. For each pair of Recipients, Social Architects choose between various monetary transfers that resemble policy reforms. These policy reforms are conditioned on the incomes of the Recipients. Social Architects' decisions across these pairs of Recipients can be used to identify the welfare weights they assign to the Recipients.

In the experiment, the key information that the Social Architects have about the Recipients is their disposable incomes. We draw on the theoretical framework of Saez and Stantcheva (2016) who show that welfare weights assigned based on the incomes of the Recipients can incorporate various ideals, such as equality of opportunity, utilitarianism, distribution based on merit vs. luck, or poverty alleviation. For example, a Social Architect guided by equality of opportunity would assign higher welfare weights to Recipients from disadvantaged backgrounds relative to other Recipients. We follow a sufficient statistics approach to estimate the welfare weights of the Social Architects. In this approach, a Social Architect's assessment of the welfare implications of a reform is a function of their welfare weights; the average welfare weights can be estimated using Social Architects' choices between various reforms without having to specify and uncover the underlying ideals that guide the Social Architect.

Our results suggest that the average welfare weights of the general population are progressive, implying that they are decreasing with the incomes of the Recipients. Progressive welfare weights indicate that, on average, the general population wants additional redistribution beyond that achieved by the current tax and transfer system.

While the average welfare weights of the Social Architects are progressive, there is considerable individual-level heterogeneity in the weights. We find that 66% of the Social Architects have progressive welfare weights, while 34% of the Social Architects have regressive welfare weights. Regressive welfare weights can be consistent with libertarian ideals, which involve limited redistribution through the tax and transfer system.

We find that Social Architects' welfare weights relate to their background characteristics. Republicans have less progressive welfare weights relative to Democrats and Independents, which is consistent with the results in the literature indicating that Republicans have weaker redistributive tastes (e.g., Fisman et al., 2017; Singhal, 2021). However, we find that, on average, Republicans have progressive welfare weights indicating that they also want additional redistribution. Social Architects with higher incomes have less progressive weights, consistent with the findings in the literature that higher-income individuals have weaker redistributive tastes (Cohn et al., 2019; Singhal, 2021).

We explore the empirical link between Social Architects' welfare weights and their preferences for government redistribution. Such an exercise highlights the value of using the welfare weights of the general population to identify socially acceptable policies. Our results suggest that Social Architects' assigned welfare weights predict their preferences for government redistribution. A benchmarking exercise reveals that the Social Architects' welfare weights predict their policy views as accurately as their stated political affiliation. We also explore the factors that predict Social Architects' preferences for redistribution. We find that while Social Architects' welfare weights predict their preferences for redistribution, their views about taxes and government play a role too. Other than Social Social Architects' welfare weights, their beliefs about the negative externalities that arise due to inequality (e.g., crime) and their beliefs about taxation on high-income individuals hurting the economy are the most predictive of their preferences for redistribution.

We compare the average welfare weights of the general population obtained from our experiment to the welfare weights implied by the income tax schedule in the U.S. computed by Hendren (2020) and by transfer policies in the U.S. computed by Hendren and Sprung-Keyser (2020). The weights implied by the tax schedule and transfer policies reflect politicians' aggregation of societal welfare weights, potentially influenced by their political economy considerations. We find that the degree of redistribution implied by the average welfare weights of the general population is about 3.7 to 5.3 times higher than

the redistribution implied by the income tax schedule, and is about 1.4 to 2 times higher than the redistribution implied by transfer policies. We explore several reasons why the general population weights are more progressive than the weights implied by the income tax schedule. Our analyses suggest that the weights implied by the income tax system are less progressive because politicians aggregating societal welfare weights underweight low-income individuals who have more progressive welfare weights and overweighting high-income individuals who have less progressive weights. This explanation is consistent with the evidence in the literature suggesting that implemented policies are often strongly influenced by the interests of higher-income individuals (e.g., Gilens, 2005; Gilens and Page, 2014).

Next, we compare the average welfare weights of the general population to the welfare weights frequently assumed in the optimal policy literature, which are inversely proportional to the incomes of the Recipients. Assuming log utilities, these weights correspond to utilitarian welfare weights. We find that the degree of redistribution implied by log-utilitarian welfare weights commonly assumed in the optimal policy literature is about 1.8 to 3.3 times higher than the redistribution implied by the average welfare weights of the general population.

To understand the implications of the welfare weights on tax policies, we calibrate the optimal labor income taxes in the U.S. using different estimates of welfare weights. We use the tax formulas derived in Saez and Stantcheva (2016). We find that the optimal marginal tax rates calibrated with the general population weights are lower than those calibrated with log-utilitarian weights, similar to those calibrated with the weights implied by transfer policies, and higher than those calibrated with the weights implied by the tax schedule.

Our paper is related to four strands of literature. The first is the experimental literature that aims to understand the factors or ideals that influence people's preferences for redistribution (e.g., Cappelen et al., 2013; Drenik and Perez-Truglia, 2018; Fong and Poutvaara, 2019; Schokkaert and Devooght, 2003; Weinzierl, 2017). These studies do not directly elicit welfare weights. Rather, they are interested in identifying if an ideal affects people's preferences for redistribution.

The second is the literature that aims to directly elicit people's welfare weights using surveys. Saez and Stantcheva (2016) elicit the welfare weights of a sample of participants recruited from the online labor market Amazon Mechanical Turk and calibrate the optimal linear income tax based on their data. Our experimental approach builds on their framework and allows us to estimate welfare weights that are applicable to a much larger set of policies, such as taxation, cash transfers, and in-kind transfers. In our paper, we also

explore the individual heterogeneity in people's assigned welfare weights, validate the welfare weights by checking if they correlate with preferences for government redistribution, and compare the general population weights to the weights implied by the tax system in the U.S. and the weights commonly used in the optimal policy literature.

The third is the literature that aims to incorporate a parsimonious set of ideals in optimal policy formulas by modifying the utility functions of the Recipients or the objective function (e.g., Fleurbaey and Maniquet, 2006; Weinzierl, 2014, 2018). For example, Weinzierl (2014) shows how the principle of equal sacrifice can be incorporated into the objective function of optimal tax formulas. In contrast, our approach does not involve modifying the utility functions of the Recipients or the objective function but instead elicits the welfare weights of the general population and plugs these weights into the standard optimal policy formulas. Our experimental results suggest that Social Architects' welfare weights are heterogeneous. Furthermore, we find that the variation in Social Architects' welfare weights can be explained by their welfarist concerns (i.e., welfare weights depend on characteristics that affect the Recipients' utilities) and non-welfarist concerns (i.e., welfare weights depend on characteristics that do affect the Recipients' utilities). These findings support our general sufficient-statistics approach of allowing Social Architects to be guided by various welfarist and non-welfarist ideals.

The fourth is the literature that identifies the social marginal welfare weights implied by the tax schedule of a country (e.g., Bourguignon and Spadaro, 2012; Hendren, 2020; Lockwood and Weinzierl, 2016; Zoutman et al., 2013) or implied by transfer policies (Hendren and Sprung-Keyser, 2020). There are several limitations to using these "inverse-optimum weights" (see Stantcheva (2016) and Lockwood and Weinzierl (2016)). First, they may not be normatively appealing if the processes that led to these weights are undesirable (e.g. if politicians are influenced by political economy considerations or lobbying). Second, they can sometimes be negative, in which case, they cannot be used in standard policy formulas that only allow positive welfare weights. Third, they are sensitive to the assumptions about the elasticity of taxable income.

Our paper proceeds as follows. In Section 2, we present the theoretical framework. The theoretical framework helps understand the experimental design that follows in Section 3. Section 4 explores the welfare weights of the general population. Section 5 compares the general population weights to the weights implied by tax and transfer policies and to the weights frequently used in the optimal policy literature. Section 6 explores the ideals that

guide the Social Architects. Finally, Section 7 presents a discussion of the paper.

2 Theoretical Framework

We present a simplified version of the model by Saez and Stantcheva (2016). Section 2.1 describes the setup of the model, Section 2.2 describes welfare weights, Section 2.3 explains how welfare weights can be identified, and 2.4 presents a discussion of our approach.

2.1 Setup

Consider a population of N *Recipients* indexed by j. A Recipient j's money-metric indirect utility function is given by $U_j = z_j - T_j(z_j) - v(z_j, \gamma_j)$ where z_j denotes Recipient j's optimal choice of income given the tax schedule, T_j denotes the taxes paid by Recipient j given the optimal choice of income, and v represents the disutility of work.[1] The disutility of work increases with income. Implicitly, this captures the idea that achieving higher income requires higher effort. The disutility of work also depends on various personal characteristics γ_j (e.g., disability status).

2.2 Welfare Weights

A *Social Architect* has a model of the world in mind. The Architect assigns *generalized social marginal welfare weight* (henceforth *welfare weight*) to Recipient j given by

$$g_j = g(c_j, \theta_j) \tag{1}$$

where $c_j = z_j - T_j(z_j)$ is the consumption of Recipient j and the vector θ_j includes a set of characteristics.[2] Some characteristics in θ_j may also be included in γ_j (e.g., disability status), while other characteristics may be excluded from γ_j (e.g., parental income). The welfare weight g_j measures the relative social value of a \$1 increase in consumption to Recipient j. Since the welfare weights are relative, they are defined up to a multiplicative constant.

The welfare weights in Equation (1) are general enough to capture various ideals such as equality of opportunity, utilitarianism, distribution based on merit versus luck, or poverty alleviation. For example, a Social Architect guided by equality of opportunity as an ideal would assign higher welfare weights to Recipients from disadvantaged backgrounds (captured by θ_j) relative to other Recipients. A utilitarian Social Architect may assign

[1] In this context, money metric-utilities imply that we remove the common concave transformation of the utilities such that all the utility gains and losses are expressed in dollar terms. However, the welfare weights estimation does not rely on this assumption.

[2] The welfare weights can additionally depend on the taxes (T_j). We assume that T_j is contained in θ_j.

welfare weights that are inversely proportional to the consumption of the Recipients. An overview of ideals that can be incorporated by this approach can be found in Saez and Stantcheva (2016).

In this theoretical framework, policies are conditioned only on the Recipients' incomes. We assume that a Social Architect reports their welfare weights averaged across all the Recipients with the same income.[3] The average welfare weights assigned to Recipient j earning z is the average of the welfare weights assigned to all individuals earning z and is given by

$$\bar{g}_j = \frac{\sum_{j:z_j=z} g_j}{h(z)}. \tag{2}$$

where $h(z)$ is the fraction of Recipients with earnings z. A set of average welfare weights assigned to the Recipients can result from one or more underlying ideals. For example, a Social Architect guided by equality of opportunity or utilitarianism may have the same average welfare weights. A Social Architect guided by equality of opportunity may assign higher welfare weights to Recipients from disadvantaged backgrounds. Since the share of individuals from disadvantaged backgrounds decreases as incomes rise, the average welfare weights also decreases as incomes rise. Utilitarian welfare weights can also result in average welfare weights decreasing with incomes. There exists a distribution of individuals from disadvantaged backgrounds that makes the average welfare weights corresponding to these two ideals identical.

2.3 Identifying Welfare Weights

Consider a setting with one Social Architect and two Recipients with incomes z_l and z_h, such that $z_h > z_l$. We consider a "small" (marginal) reform that is implemented given the current tax system. The reform $\epsilon R = (\epsilon R_l, -\epsilon R_h)$ takes ϵR_h from the higher-income Recipient and gives ϵR_l to the lower-income Recipient. The reform mechanically changes the disposable income of the Recipients by $(\epsilon R_l, -\epsilon R_h)$.[4] The Social Welfare function is given as follows.

[3] When policies are conditioned jointly on incomes and other observable characteristics ("tags"), there are two ways to evaluate policies. The first approach involves averaging the weights up to the level of the policy, i.e., jointly on incomes and other observable characteristics. The second approach is to ignore the tags when comparing policies and averaging the weights up to the incomes of the Recipients. For an implementation of the latter approach, see Hendren and Sprung-Keyser (2020).

[4] Because the Recipients are optimizing, the reform does not affect their pre-tax income due to the envelope condition.

$$
\begin{aligned}
W &= \bar{g}_l \cdot U_l + \bar{g}_h \cdot U_h \\
&= \bar{g}_l \cdot (z_l - T_l(z_l) + \epsilon R_l - v(z_l, \gamma_l)) + \bar{g}_h \cdot (z_h - T_h(z_h) - \epsilon R_h - v(z_h, \gamma_h)). \qquad (3)
\end{aligned}
$$

The social welfare function weights the utilities of the Recipients by the average welfare weights assigned to the Recipients. Taking the first derivative of Equation (3) with respect to ϵ, we get

$$\Delta W = \bar{g}_l \cdot R_l + \bar{g}_h \cdot (-R_h). \qquad (4)$$

This equation highlights that the weighted reform amounts, weighted by the average welfare weights, have a first-order effect on welfare. To identify the average welfare weights $\bar{g}_l$ and $\bar{g}_h$, we need to identify two reforms, R and R', that the Social Architect is indifferent between. Setting $\Delta W(R) = \Delta W(R')$ allows to recover the average welfare weights $\bar{g}_l$ and $\bar{g}_h$.

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

2.4 Discussion of the Approach

There are three important aspects to our approach that warrant a discussion. First, a Social Architect does not maximize any social welfare objective. Instead, the Social Architect assigns welfare weights to individuals in society given the tax and transfer system and reports their average welfare weights. These average welfare weights can be used to determine the Social Architect's desirability of a small reform using Equation (4). The social welfare function is defined above only to derive Equation (4).

Second, our approach to estimating welfare weights is based on the sufficient statistics approach. In this approach, a Social Architect's assessment of the welfare implications of a reform (in Equation (4)) is a function of their average welfare weights, which are "high-level" estimable statistics rather than deep primitives. As discussed above, we can estimate these average welfare weights based on Social Architects' choices between various reforms without having to specify and uncover the underlying ideals (primitives) that guide the Social Architect. Even though a Social Architect's average welfare weights can be consistent with multiple underlying ideals (e.g., equality of opportunity or utilitarianism), these ideals have the same welfare implications.

Third, a key concern in our approach is that the welfare weights assigned by a Social Architect 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, a Social Architect may have misperceptions about characteristics that do not enter θ in Equation (1), implying that these misperceptions do not affect their welfare weights. For example, a Social Architect guided only by equality of opportunity may have a misperception about the share of disabled individuals, but this misperception would not affect their welfare weights. Second, the welfare weights averaged over many Social Architects may be unaffected by misperceptions if these misperceptions across thee Social Architects sum to zero. In the following sections, we present the average welfare weights of Social Architects and discuss how these average weights can be used to evaluate policies.

3 Experimental Design

Section 3.1 explains the key features of the design required to elicit Social Architects' welfare weights. Section 3.2 describes the experimental procedures. Section 3.3 describes the treatments fielded in the two waves of data collection. Section 3.4 describes the additional questions fielded in the two waves. Section 3.5 describes the data collection. Section 3.6 describes the summary statistics. Finally, Section 3.7 describes the pre-registration. The full set of instructions for Wave 1 can be found in Appendix G and for Wave 2 in Appendix H.

3.1 Eliciting Welfare Weights

Participants in our study are either in the role of a "Social Architect" or "Recipient." The Recipients are passive subjects who receive money based on the Social Architects' decisions. A Social Architect faces two Recipients with incomes z_l and z_h, such that $z_h > z_l$. The Social Architect is informed that the Recipients receive a \$1500 endowment.

We are interested in eliciting the average welfare weights assigned by Social Architects to the two Recipients. For the sake of brevity, we omit the word "average" and refer to $\bar{g}_l, \bar{g}_h$ as welfare weights. As indicated in Section 2.3, to identify the welfare weights, we need to identify two reforms, R and R', that the Social Architect is indifferent between.

A Social Architect is asked to choose between a "Constant Reform" (\$500, $-$\$500) and various "Variable Reforms" of the type ($\$pt, -\$t$); the reforms ($\$pt, -\$t$) take \$t from the higher-income Recipient and give \$pt to the lower-income Recipient. We identify the reform ($\$pt, -\$t$) that makes a Social Architect indifferent between ($\$pt, -\$t$) and (\$500, $-$\$500).[5] Setting $\Delta W((\$500, \text{-}\$500)) = \Delta W((\$pt, \text{-}\$t))$ in Equation (4), we get

[5] In principle, Social Architects can be asked to choose between a reform ($\$pt, -\$t$) and no reform (\$0, \$0).

$$\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l} = \frac{pt - 500}{t - 500}. \tag{5}$$

Equation (5) shows that the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient is proportional to the parameter p. A smaller p implies that the Social Architect needs to give less to the lower-income Recipient to be indifferent between ($\$pt$, $-\$t$) and ($\$500$, $-\$500$), implying a relatively higher weight on the lower-income Recipient. When $p < 1$, the welfare weight assigned to Recipient h is lower than that assigned to Recipient l, implying progressive welfare weights. If $p > 1$, Social Architects have regressive weights. If $p = 1$, Social Architects assign the same weight to both Recipients.

Table 1 presents fifteen pairs of Variable Reforms that can be used to elicit Social Architects' welfare weights. The Variable Reforms are arranged in rows, with lower row numbers indicating larger amounts taken from the higher-income Recipient, up to their entire endowment ($1500), and smaller amounts given to the lower-income Recipient. Thus, lower row numbers correspond to Variable Reforms with lower values of p. In Row 8, the Variable Reform ($1000, $-\$1000$) implies $p = 1$. Variable Reforms above Row 8 have an implied $p < 1$, while those below Row 8 have an implied $p > 1$. The Variable Reforms below Row 8 mirror the ones above.

Table 1 is a multiple-price list. We assume that Social Architects would typically choose a Constant Reform starting from Row 1 before switching to a Variable Reform. We identify the reform ($\$pt$, $-\$t$) that makes them indifferent between ($\$pt$, $-\$t$) and ($\$500$, $-\$500$) by computing the mid-point of the Variable Reform that Social Architects switch to and the Variable Reform in the previous row. For those who switch in the first (last) row, we compute the mid-point of the Variable Reform in the first (last) row and the bound on possible reforms.[6] Social Architects who switch earlier have a lower implied p because lower row numbers correspond to lower values of p. We identify the welfare weights by plugging the mid-point of the Variable Reforms into Equation (5).[7]

However, to minimize Social Architects' susceptibility to status-quo bias, which entails choosing the status-quo ($\$0, \$0$), we avoided this approach.

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

[7] For instance, if a Social Architect prefers a Constant reform in Rows 1-6 and switches to the Variable

Table 1: Set of Possible Questions in Each Decision Screen

Row	Constant Reform	Variable Reform	p
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 a set of fifteen Variable Reforms. A Variable Reform ($\$pt, -\$t$) takes away \$t from the higher-income Recipient in the pair and gives \$pt to the lower-income Recipient in the pair. p refers to the ratio of the Variable Reform amount given to the lower-income Recipient by the Variable Reform amount taken from the higher-income Recipient.

Higher row numbers in Table 1 correspond to less progressive Variable Reforms, indicated by higher values of p. However, these reforms are also more efficient since they allocate larger amounts to lower-income Recipients while taking smaller amounts from higher-income Recipients. Social Architects who prioritize efficiency would switch later in Table 1, assigning more regressive weights. In our experimental setting, efficiency concerns may not play a big role since Social Architects may not have preferences over money in the hands of the experimenter. Furthermore, there is evidence in other contexts documenting that efficiency concerns do not play a large role in people's redistributive

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 (5), we get $\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l} = \frac{pt-500}{t-500} = \frac{912.5-500}{1087.5-500} = 0.702$. This implies that the relative weight assigned to the higher-income Recipient is 0.702 times the weight assigned to the lower-income Recipient.

decisions (Cappelen et al., 2013; Stantcheva, 2020).

3.2 Experimental Procedures

Information about Recipients

Social Architects learn 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 disposable incomes of the seven Recipients and are informed that these incomes are accrued after all taxes and transfers. We provide Social Architects with information about the disposable incomes of the Recipients because it allows us to identify the Social Architects' welfare weights given the current tax and transfer system. Finally, we inform the Social Architects that the Recipients receive a $1500 endowment.

The Recipients' incomes span the income distribution of the U.S., covering the various tax brackets. Figure 1 plots the incomes of the seven Recipients (dots) against the disposable income distribution (line) in the U.S. in 2019.

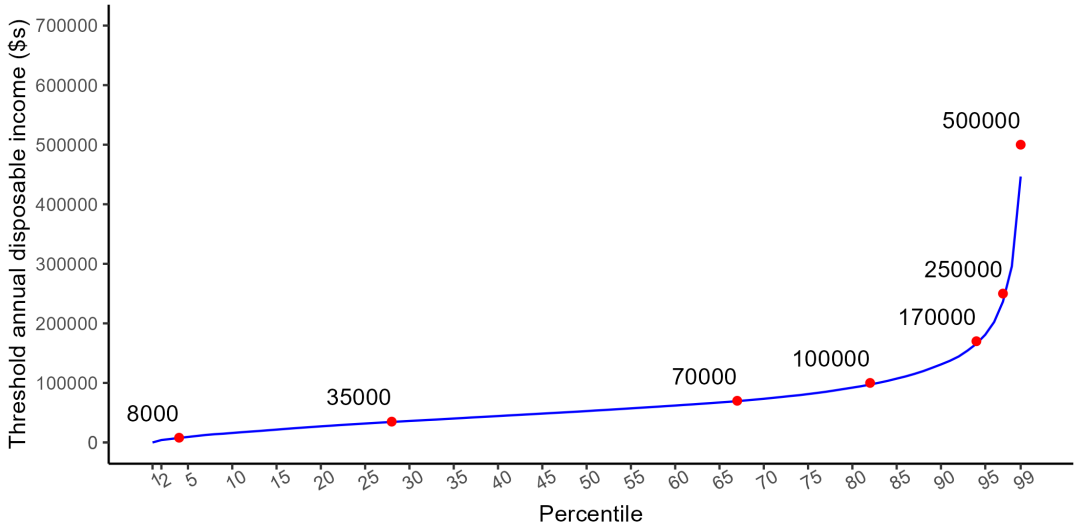


Figure 1: Disposable Incomes of the Seven Recipients

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 income required for an individual to fall in the corresponding income percentile. 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.

Decision Screens

Social Architects encounter six "decision screens," facing a pair of Recipients on each screen. Table 2 displays the incomes of the Recipients in each decision screen. For half the

participants, the order of the decision screens is reversed.

Table 2: Pairs of Recipients in the Six 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

We are interested in identifying the welfare weights that the Social Architects assign to the two Recipients in each decision screen. Since the Recipient earning \$70,000 in common across the six decision screens, Social Architects' decisions across the six decision screens allow us to identify the (relative) welfare weights assigned to each of the seven Recipients.

In each decision screen, we want to identify, based on the questions indicated in Table 1, the Variable Reform ($\$pt, -\$t$) that the Social Architects switch to from a Constant Reform (\$500, −\$500). To do so, we use the "staircase method," which selects four of the fifteen questions presented in Table 1 in adaptively. The first question that the Social Architects face is the one in Row 8: (\$500, −\$500) vs. (\$1000, −\$1000). The second, third, and fourth questions that the Social Architects face depend on their choices in the first, second, and third questions, respectively.[8] Figure 2 presents a graphical representation of the questions faced by Social Architects.

The staircase method enforces that the Social Architects have a unique switch point and can only switch from a Constant Reform to a Variable Reform. It also ensures that a Social Architect who chooses a Variable Reform in the first question (implying progressive welfare weights) subsequently only faces questions from the top half of the table, which are used to identify the progressivity of the weights. Similarly, a Social Architect who chooses a Constant Reform in the first question (implying regressive welfare weights) subsequently only faces questions from the bottom half of the table, which are used to identify the regressivity of the weights. This feature ensures that if Social Architects answer the first question correctly, the qualitative assessment of their welfare weights (whether progressive or regressive) is free from measurement error.

Figure A1 in Appendix Section C.1 presents a screenshot of one of the questions used in the experiment. 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

[8]The staircase method has several advantages. First, this method is easy to explain to the participants, especially to a nationally representative sample. Falk et al. (2018) use the staircase method in nationally representative samples across the world, highlighting its simplicity. Second, this method allows us to get more accurate weights with fewer questions compared to using a multiple-price list. In Appendix F.1, we provide a proof of the incentive compatibility of the staircase method.

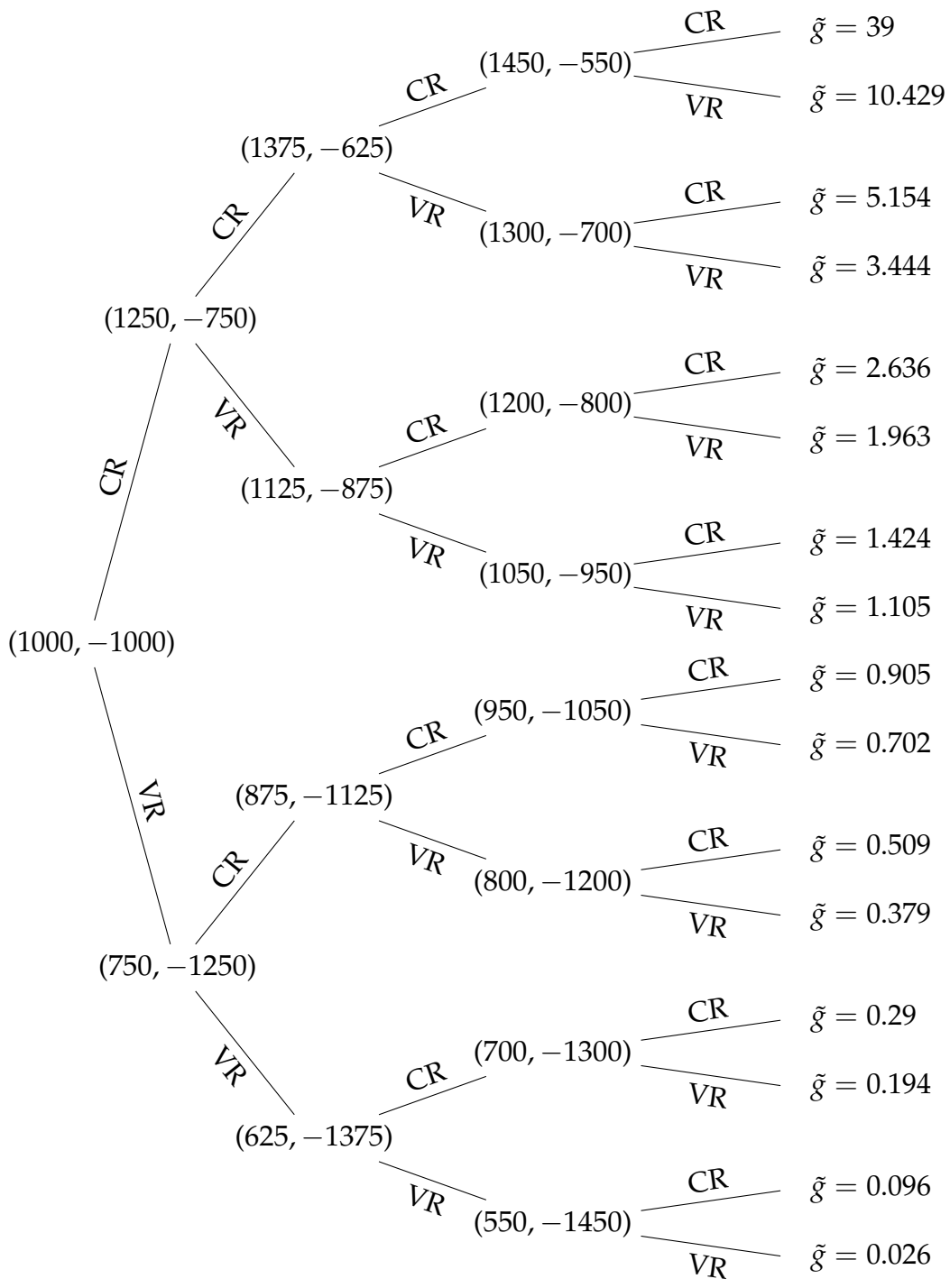


Figure 2: Variable Reforms Selected by the Staircase Method

Notes: "VR" and "CR" indicate that the Variable Reform and the Constant Reform was chosen in the previous node, respectively. The Constant 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.

income." Finally, we refer to the $1500 endowment as an “initial bonus."

Incentives

We inform the Social Architects that we will randomly select one of them in the study (including all waves of data collection). For the randomly selected Social Architect, one randomly selected question in one randomly selected decision screen will be implemented. At the top of each question, we remind Social Architects to consider each question carefully since one randomly selected question will have real consequences for two Recipients. At the end of the study, two Recipients would receive ($1500+pt, $1500-t) or ($1500+500, $1500-500), depending on the choice of the selected Social Architect on the selected question.

3.3 Treatments and Waves

We implemented eight treatments across two waves of data collection to test the robustness of the elicited welfare weights with respect to changes in the experimental design. Table 3 presents an overview of the treatments.

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 (explained above) 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 Social Architects decide between the reforms ($500, $-\$500$) and ($\$pt$, $-\$t$). In Treatments Gain, there is no additional endowment, and the Social Architects decide between ($2000, $1000) and ($\$1500 + pt, \$1500 - t$). The framing of the reform should not affect Social Architects' welfare weights since the welfare weights depend on the Recipients' consumption, which is the same in the two treatments. However, if Social Architects are influenced by loss aversion, they will be less likely to choose the Variable Reform in Treatments Loss, and consequently, assign less progressive welfare weights.

The second dimension helps test if Social Architects' welfare weights are sensitive to the income of the Recipient common across the decision screens. In Treatments 70K (explained above), 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.

In Wave 2 of data collection, Social Architects are randomly assigned to one of four treatments. Treatment Base in Wave 2 is similar to Treatments Loss $\times$ 70K in Wave 1. We include this treatment to test if there are differences across waves. The only difference between the two treatments is that in Treatment Base, we mentioned that the “after-tax"

incomes of the Recipients were accrued after all taxes and transfers, which allows us to identify Social Architects' welfare weights at the margin.

Treatment Hypothetical is identical to Treatment Base except that Social Architects make decisions regarding seven hypothetical Recipients and are informed that their decisions are hypothetical. This treatment allows us to test if Social Architects' welfare weights are sensitive to the existence of real stakes (albeit small stakes). If the welfare weights from the two treatments are similar, then future research can use hypothetical decisions to elicit welfare weights since they are easier to implement. However, if the welfare weights from the two treatments are not similar, then we defer to Treatment Base since the presence of real stakes is likely to lead to more reliable welfare weights.

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 themselves. To examine the role of self-interest motives, we include Treatment Self-Interest, in which Social Architects can potentially receive a bonus from their own choices. This treatment differs from Treatment Base 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 $10,000, they would replace the first Recipient whose income bracket is "$22,000 and below."

We compare Social Architects in Treatment Self-Interest to Treatment Brackets. Treatment Brackets 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 Brackets cannot be affected by their own decisions.

3.4 Additional Questions

In this section, we present an overview of the additional questions that Social Architects are asked to answer. These questions can be found in Appendix Sections G and H.

Wave 1

We elicit Social Architects' views on the taxes levied on those in the top-income tax category and whether the government should reduce income differences between the rich and the poor. We use these two validated questions that are frequently used in the literature to test whether Social Architects' welfare weights predict their preferences for government

Table 3: Overview of Treatments

Wave 1		
Treatment	Framing of Reforms	Income of Common Recipient
Loss × 70K	Loss	70K
Loss × 500K	Loss	500K
Gain × 70K	Gain	70K
Gain × 500K	Gain	500K
Wave 2		
Treatment	Framing of Reforms	Income of Common Recipient
Base	Loss	70K
Hypothetical	Loss	70K
Self-interest	Loss	70K
Brackets	Loss	70K

redistribution. Such an exercise highlights the value of using the welfare weights of the general population to identify socially acceptable policies.

Wave 2

We elicit Social Architects' preferences for government redistribution by asking them to consider the current incomes of individuals in society after all taxes and transfers and asking them if they would like to redistribute incomes further. 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 their beliefs about the status quo beyond which redistribution should occur. Second, the answer options include transfers from the poor to the rich, which is useful to capture ideals such as libertarianism. We use this question to test whether Social Architects' welfare weights predict their preferences for government redistribution.

We ask Social Architects about their beliefs regarding whether high-income individuals and low-income individuals deserve and need their current income. We use these two questions to test whether Social Architects' assigned welfare weights depend on their beliefs about the Recipients' needs or on factors orthogonal to the Recipients' needs. Social Architects' judgments about whether individuals deserve their incomes are guided by "non-welfarist" motives, i.e., their welfare weights depend on characteristics that do not directly affect Recipients' utility functions (e.g., Recipients' parental income). Testing whether Social Architects' are guided by non-welfarist concerns is important because the value of the approach developed by Saez and Stantcheva (2016) lies in incorporating non-welfarist motives in welfare weights.

We ask Social Architects how confident they are that their choices in the welfare weights elicitation task reflect what they really think. We ask this question to learn if the Social Architects are confident in their choices and to explore the welfare weights of those who report having low confidence.

Finally, to test whether Social Architects' welfare weights are influenced by their beliefs about the source of income, we elicit their beliefs about whether high-income individuals are rich due to luck or effort.

3.5 Data Collection

In Wave 1 of data collection, we recruited participants in the role of Social Architects from the data collection provider Lucid.[9] The collected sample includes participants recruited based on quotas for gender, age, education, individual income, and region. The quotas are designed to match the sample to the population of the U.S. Participants first answer questions about their demographics and political affiliation. Next, participants face 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 face two comprehension checks. Participants who answer either of the two comprehension checks 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.[10]

In Wave 2 of data collection, we recruited participants in the role of Social Architects from the data collection provider Prolific.[11] In Wave 2, we used Prolfic instead of Lucid because Prolific allows us to pay participants a bonus, a feature required in Treatment Self-Interest and not available in Lucid. 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 approximately a week. Our

[9] Lucid is commonly used in the literature (e.g., Haaland and Roth, 2023; Haaland et al., 2021).

[10] We recruited a total of 6,735 participants. We drop participants who attempted the survey more than once (0.09%), who do not consent to participate in the study (2.5%), who do not fit into one of the demographic quotas or do not reside in the U.S. (28.1%), who drop out before the attention check (6.4%), who fail the attention check (28.2%), who drop out before the comprehension check (16.5%), who fail the comprehension check (21.3%), who drop out after passing the comprehension check (5.8%). The share of participants who pass the attention check but drop out before the comprehension checks is not different across the four treatments ($F = 0.5228, p = 0.66$). The share of participants that pass the comprehension checks is not different across the four treatments ($F = 1.763, p = 0.1522$).

[11] Prolific has been used in several recent studies (e.g., Bursztyn et al., 2023; Enke et al., 2022)

final sample includes 1992 participants.[12]

3.6 Summary Statistics

Table 4 presents the average characteristics of our sample relative to the characteristics of the population of the U.S. in 2019. The sample characteristics in Wave 1 closely match the population characteristics because we implemented quotas based on the average population characteristics. However, 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. The sample characteristics in Wave 2 also broadly match the population characteristics, but there are notable differences. 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. In the following sections, we report analyses that are weighted using sampling weights, which are constructed to ensure that the weighted averages of our sample characteristics match those of the population.

Table A1 and Table A2 in Appendix 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 exploring treatment effects, we account for these imbalances by weighting each treatment using sampling weights to match the population.

3.7 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 that are discussed in Appendix Section B.

4 Understanding Welfare Weights

4.1 Distribution of Welfare Weights

Figure 3 presents the distribution of the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient, across the six pairs of Recipients (decision

[12]We recruited a total of 2,313 participants. We drop participants who attempted the survey more than once (0.3%), who do not consent to participate in the study (0.04%), who drop out before the attention check (1.7%), who fail the attention check (2%), who drop out before the comprehension check (3.1%), who fail the comprehension check (5.9%), who drop out after passing the comprehension check (1.6%). The share of participants who pass the attention check but drop out before the comprehension checks is not different across the four treatments ($F = 0.7024$, $p = 0.5506$). The share of participants that pass the comprehension checks is not different across the four treatments ($F = 0.168$, $p = 0.918$).

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 is restricted to those 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.

screens). The relative weight $\tilde{g}$ is calculated using Equation (5). Each sub-figure includes 16 bars, each of which corresponds to a row in Table 1 where the Social Architects switch from a Constant Reform to a Variable Reform. The last bar corresponds to those who never switch. The figure uses the data of Social Architects in all treatments in both waves where the Recipient common across the pairs of Recipients has an income of \$70,000.

We observe considerable heterogeneity in welfare weights across the pairs of Recipients, comprising both progressive welfare weights ($\tilde{g} < 1$), where welfare weights decrease with Recipients' incomes, and regressive welfare weights ($\tilde{g} > 1$), where welfare weights increase with Recipients' incomes. This heterogeneity also extends within these two groups. Overall, the distribution of welfare weights across the pairs of Recipients is

skewed towards progressivity, suggesting that the welfare weights assigned to the seven Recipients are, on average, progressive.

We also observe bunching at three values of $\tilde{g}$ across the six pairs of Recipients. The first is at $\tilde{g} = 0.03$, representing the most progressive welfare weights. These weights are observed when a Social Architect, facing a pair of Recipients, selects the Variable Reform in all four questions. Selecting the Variable Reform in all four questions implies that this Social Architect is willing to give the smallest amount to the lower-income Recipient and take the largest possible amount from the higher-income Recipient. The second is at $\tilde{g} = 0.09$, representing the least progressive among the progressive welfare weights. These weights occur when a Social Architect selects the Variable Reform in the first question and the Constant Reform in the subsequent three questions. Finally, the third is at $\tilde{g} = 39$, which implies the most regressive welfare weights. These weights are observed when a Social Architect chooses the Constant Reform in all four questions.[13]

Figure 3 in Appendix Section C.1 presents the distribution of $\tilde{g}$ using treatments in the first wave where the Recipient common across the pairs of Recipients has an income of \$500,000. In these treatments, we observe a similar shift in the distribution towards progressivity and similar bunching.

4.2 Average Welfare Weights

In this section, we explore the average welfare weights assigned by the Social Architects in all treatments and waves. We discuss issues in the aggregation of welfare weights in Section 7.

We estimate the progressivity of the average welfare weights by estimating the parameter ν that makes the function y^{ν} a best fit of the welfare weights, where y is the incomes of the Recipients, and ν is a parameter that governs the progressivity of the welfare weights. The parameter ν can be interpreted as the elasticity of the welfare weights with respect to the incomes of the Recipients. Using the parametric function y^{ν} allows the welfare weights to be portable across contexts. This function is commonly used in the optimal policy literature to characterize welfare weights (e.g., Allcott et al., 2019; Saez, 2002). Negative values of ν indicate progressive weights, while positive values of ν indicate regressive weights. A value of $\nu = |0.25|$ corresponds to weak redistributive preferences, a value of $\nu = |1|$ corresponds to fairly strong redistributive preferences, and a value of $\nu = |4|$ corresponds to extremely strong redistributive preferences (Allcott et al., 2019; Saez, 2002). Based on our experimental design, the range of the possible values of ν is $(-2.25, 2.25)$. We find that only 15% of the Social Architects have an individual-level ν that is -2.25 (10%) or

[13]Social Architects with regressive weights may be guided by ideals such as libertarianism, which opposes redistribution.

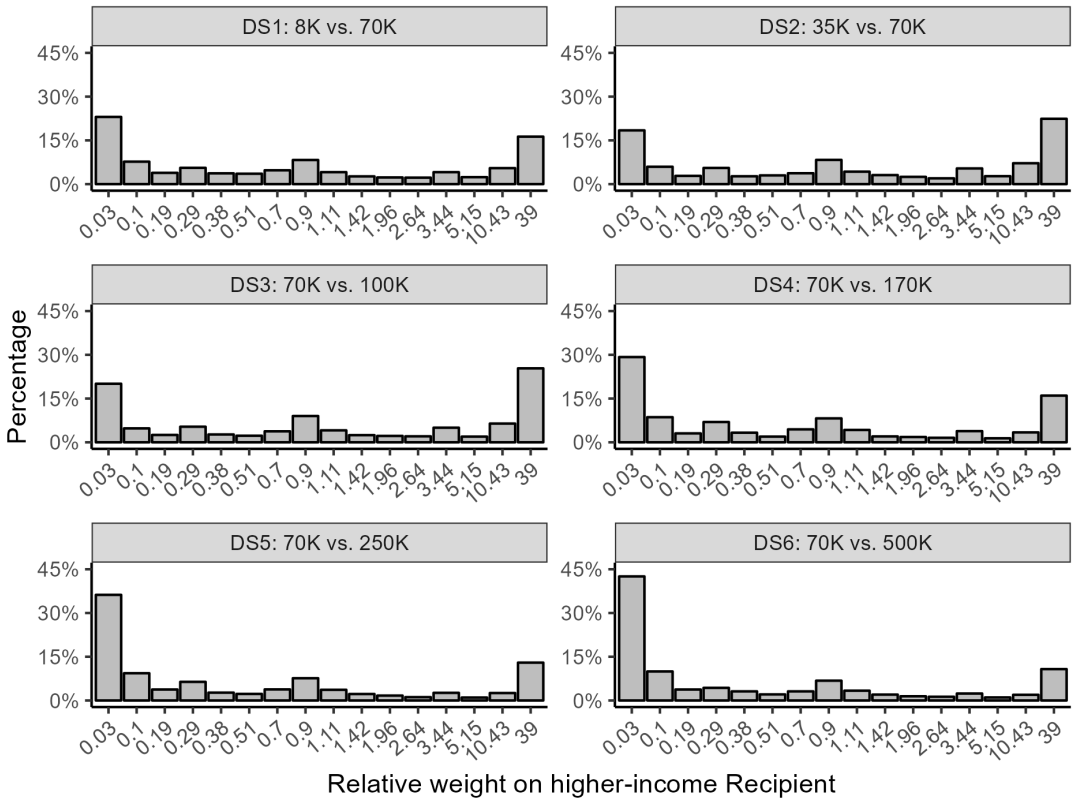


Figure 3: Distribution of $\tilde{g}$ Across Decision Screens

Notes: The figure presents the distribution of the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient across the six decision screen ("DS"). The relative weight $\tilde{g}$ is calculated using Equation (5). Social Architects are presented with a different pair of Recipients in each decision screen. For example, in the first decision screen, Social Architects face Recipients earning \$8,000 and \$70,000. For half the participants, the order of the decision screens was reversed. The sample is restricted to treatments where the Recipient common across decision screens has an income of \$70,000.

+2.25 (5%), suggesting that the limited range of ν does not severely affect our inferences. To estimate the parameter ν, we estimate the following linear regression.

$$log(g(R_j)_i) = \beta_0 + \nu log(recipient\ income_j) + \epsilon_{ij} \tag{6}$$

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 demographics in the sample match the average demographics in the population.[14]

[14]The average population characteristics are taken from Column (1) in Table 4. The procedure creates sampling weights, sometimes referred to as "raking" weights, based on a characteristic only if the sample

Our findings indicate that the estimated value of ν is -0.34. This estimate is close to 0.25, suggesting that, on average, Social Architects have "weak" redistributive preferences. This elasticity estimate implies that a Social Architect is indifferent between giving a dollar to a Recipient and giving 66 cents to a Recipient earning twice as much. If the estimated ν were close to zero, it would mean that Social Architects are indifferent between giving a dollar to the rich and giving a dollar to the poor, implying that they are satisfied with the redistribution achieved by the current tax and transfer system in the U.S. However, we find that the estimated ν is less than zero, suggesting that, on average, the general population wants additional redistribution beyond that achieved by the current tax and transfer system. This leads to the following result.

Result 1. *On average, the general population in the U.S. want additional redistribution beyond that achieved by the current tax and transfer system.*

4.3 Treatment Effects

In this section, we compare the average welfare weights assigned by the Social Architects across treatments by estimating the following linear regression.

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

where $log(g(R_j)_i)$ is the natural logarithm of the welfare weight assigned by Social Architect i to Recipient j and $\textit{recipient income}_j$ is the disposable income of Recipient j. The variables $x^1, ..., x^n$ (shortened to X) include a set of treatment dummies from Wave 1 and Wave 2. Since Social Architects' characteristics are not balanced across treatments, we also present weighted regressions using sampling weights constructed such that the average demographics in each treatment match the average demographics in the population.

Figure 4 plots the coefficient estimates $\nu_0, .., \nu_n$. The coefficient estimate of $\textit{log(recipient income)}_j$ indicates the estimated elasticity of the weights with respect to the incomes of the Recipients in Treatment Loss $\times$ 70K (the base category). We find that the estimated elasticity of Social Architects' welfare weights in Treatment Loss $\times$ 70K is -0.37.

In Treatment Loss $\times$ 70K, participants face reforms that take money away from the higher-income Recipient and give money to the lower-income Recipient. In contrast, in

average is more than five percentage points away from the population average. We do not set an upper bound on the sampling weight for any observation.

Treatment Gain $\times$ 70K, the reforms are framed as a gain to both Recipients. We observe that Social Architects have more progressive welfare weights in Treatment Gain $\times$ 70K (ν is 0.14 lower) relative to Treatment Loss $\times$ 70K. In Treatments Loss, Social Architects may be reluctant to take money away from the higher-income Recipients due to loss-aversion (Charité et al., 2015). A higher reluctance to take money away from the higher-income Recipient would lead to less progressive welfare weights. In the treatments with a gain framing, loss-aversion is likely to play a smaller role since the reforms are framed as a gain to both the Recipient in the pair. While the estimated treatment difference is statistically significant, the difference is modest, indicating that the welfare weights estimation is not very sensitive to the framing of the reform.

We compare Treatment Loss $\times$ 70K to Treatment Loss $\times$ 500K. The Recipient common across the decision screens has an income of \$70,000 in Treatment Loss $\times$ 70K and an income of \$500,000 in Treatment Loss $\times$ 500K. We find that Social Architects have less progressive welfare weights in Treatment Loss $\times$ 500K (ν is 0.27 higher). A part of the treatment difference is a mechanical effect. In Treatment Loss $\times$ 500K, a Recipient who always chooses the Variable Reform in each question will assign the same welfare 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. Although there is also a mechanical flattening of the weights in Treatment Loss $\times$ 70K, the extent of the flattening is lower because the Social Architects' 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. If Social Architects choose the Variable Reform in every question, the individual level elasticity of the weights with respect to the incomes of the Recipients 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.

Treatment Loss $\times$ 70K was implemented in Wave 1 of data collection, while Treatment Base was implemented in Wave 2 of data collection. These two treatments are very similar.[15] We find that Social Architects are more progressive (ν is 0.15 lower) in Treatment Base relative to treatment Loss $\times$ 70K. However, the estimated difference is modest. The patterns in the welfare weights across the two treatments is presented in Table A3 in Appendix Section C.2.

Next, we examine whether the presence of real stakes affects Social Architects' welfare weights. To explore the role of real stakes, we compare Treatment Base to Treatment

[15]The only difference between the two treatments is that in Treatment Base, we mentioned that the "after-tax" incomes of the Recipients were accrued after all taxes and transfers.

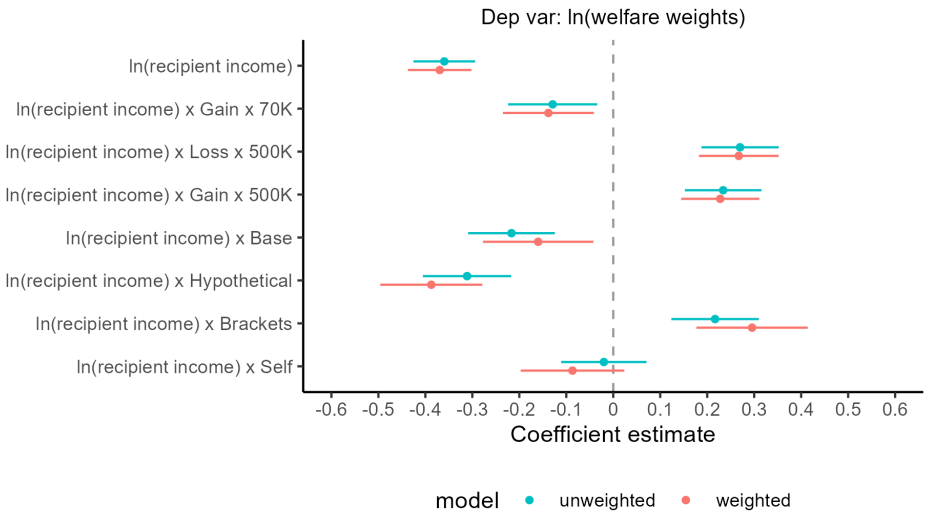


Figure 4: Welfare Weights and Treatments

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. We do not present the main effects of the treatment dummies. The weighted regression weights each treatment using sampling weights. Error bars are computed using HC3 standard errors.

Hypothetical. In Treatment Hypothetical, Social Architects faced hypothetical Recipients and made hypothetical choices. We find that participants' welfare weights are more progressive in Treatment Hypothetical (ν is 0.23 higher) relative to Treatment Base. We explore whether Social Architects are more progressive in Treatment Hypothetical because they are less attentive in the survey. The results are presented in Appendix Section D.2. We do not find evidence that participants in Treatment Hypothetical are less attentive in the survey.

Finally, we explore the role of self-interest motives by comparing Social Architects in Treatments Brackets and Self-Interest. In Treatment Self-Interest, each Social Architect replaces the Recipient whose income bracket contains their own income. Thus, each Social Architect can potentially be affected by their own choices. We find that participants in Treatment Self-interest have more progressive weights than participants in Treatment Brackets, indicating that self-interest motives do play a role. In Appendix Section D.2, we explore the role of self-interest separately for each income group. This leads to the following results.

Result 2. *Social Architects are more progressive in (i) Treatments Gain relative to Loss, (ii) Treatments 70K relative to 500K, (iii) Treatment Base relative to Treatment Loss $\times$ 70K, (iv) Treatment Hypothetical relative to Treatment Base, (v) Treatment Self-Interest relative to Treatment Brackets.*

4.4 Heterogeneity in Welfare Weights

We explore the individual heterogeneity in the progressivity of the welfare weights assigned by Social Architects. In particular, we explore the heterogeneity in the individual-level elasticity of the welfare weights with respect to the incomes of the Recipients by estimating the following regression for each Social Architect.[16]

$$log(g(R_j)) = \beta_0 + \nu log(\textit{recipient income}_j) + \epsilon_j. \tag{8}$$

The variable $g(R_j)$ is the welfare weight assigned by Social Architect i to Recipient j, $\textit{recipient income}_j$ is the income of Recipient j, and ν is the elasticity of a Social Architect's welfare weights with respect to the incomes of the Recipients. We estimate these regressions using Social Architects from all the treatments in both waves.

Figure A3 in Appendix C.1 shows the cumulative distribution function of Social Architects' estimated elasticity parameters. Around 65% of the Social Architects have progressive welfare weights ($\nu < 1$), while 35% of the Social Architects have regressive welfare weights ($\nu > 1$). There is considerable heterogeneity in the estimated elasticity parameters within the two groups. 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 the Social Architects' welfare weights.

We also explore the heterogeneity in the slope of the Social Architect's welfare weights with respect to the index of the Recipients. The slope of the weights corresponds to the parameter estimate $\hat{\beta_1}$ obtained by estimating the following 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 range from \$8,000 through \$500,000. The cumulative distribution function of Social Architects' estimated slope parameters is presented in Figure A4 in Appendix C.1. We observe similar heterogeneity in the slope of the weights as that in the elasticity of the weights.

4.5 Welfare Weights and Background Characteristics

We examine how Social Architects' welfare weights vary with their background charac-

[16] 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 do not deserve additional money because they are "lazy" (Drenik and Perez-Truglia, 2018).

teristics. We estimate a version of Equation (7), where the explanatory variables include 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. Figure 5 presents the results.

We find that Republicans assign less progressive weights (ν is 0.24 higher) relative to Democrats and Independents, which is consistent with previous studies suggesting that Republicans tend to have weaker redistributive preferences (Fisman et al., 2017; Singhal, 2021). However, 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). This result is consistent with findings in the literature indicating that higher-income individuals have weaker redistributive preferences (Cohn et al., 2019; Singhal, 2021). We further explore how Social Architects' own income affects their welfare weights. In particular, we investigate whether Social Architects assign a higher weight to Recipients with similar incomes to their own compared to other Recipients. We present the results in Appendix Section D.1. Our results suggest that Social Architects assign a higher weight to Recipients with incomes similar to their own, implying that self-interest motives play a role. This effect is larger for Social Architects with lower incomes.

Social Architects with higher levels of education have more progressive weights (ν is 0.08 lower), while older Architects have less progressive weights (with ν being 0.03 higher). The sex of the Social Architects does not predict their welfare weights.

As a robustness check, we estimate five separate regressions, with each regression only considering one characteristic of the Social Architects. These regressions also exclude treatment dummies. Figure A5 in Appendix Section C.1 presents the results in Models s2 to s6. We find very similar results. In this figure, we also present the results from Figure 5 in Model s1 as a benchmark.

As another robustness check, we estimate linear regressions in which the dependent variable is the individual-level elasticity of the Social Architects' weights with respect to the incomes of the Recipients, and the explanatory variables include Social Architects' background characteristics. Table A5 in Appendix Section C.2 presents the results. The results from these regressions are similar to those found in Figure 5. We now have the

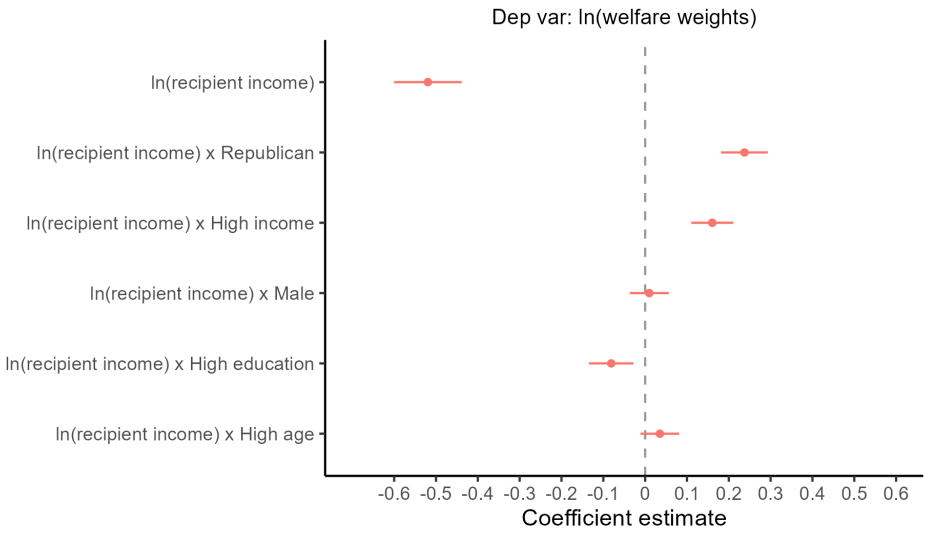


Figure 5: Social Architects' Welfare Weights and Characteristics

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. With the exception of the main effect of the log of the incomes of the Recipients, we do not present the main effects of the other variables. The regressions are weighted using sampling weights. Error bars are computed using HC3 standard errors.

following result.

Result 3. *Social Architects with higher incomes, lower education levels, and older ages, along with those identifying as Republicans, assign less progressive welfare weights.*

4.6 Welfare Weights and Preferences for Redistribution

4.6.1 Do Welfare Weights Predict Preferences for Redistribution?

We explore the empirical link between Social Architects' welfare weights and their preferences for government redistribution. Such an 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' preferences for redistribution using a survey measure. The question asks Social Architects to consider the current incomes of individuals in society after all taxes and transfers and asks them if they would like to redistribute incomes further. 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 transfers from the poor to the rich, which is useful to capture ideals such as libertarianism. We present the frequencies of responses to this question in Figure A8 in Appendix Section C.1.

We estimate the correlation between Social Architects' elasticity of the weights and their preferences for redistribution. The elasticity of Social Architects' weights is computed using Equation (8). We find that the correlation between Social Architects' elasticity of the weights and their preferences for redistribution is -0.36, which is significant at the 1% level. This result suggests that Social Architects with more progressive welfare weights have stronger preferences for progressive redistribution.

To understand how big a role welfare weights play in predicting Social Architects' preferences for redistribution, we benchmark the predictive power of the elasticity of Social Architects' welfare weights against their stated political affiliation. Political affiliation has been identified as a consistent predictor of people's preferences for redistribution (e.g., Stantcheva, 2020). We regress Social Architects' preferences 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.[17] The lower the RMSE of a specification, the higher the predictive power of that specification.

Table 5 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 preferences 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 do the benchmarking exercise based on two validated measures of preferences for redistribution that are used in the literature. The results are presented in Appendix Section D.3. This leads to the following result.

Result 4. *Social Architects' preferences for redistribution can be predicted with similar accuracy using either their stated political affiliation or their assigned welfare weights.*

Looking at Rows 5 and 6 in Table 5, we find that in a specification that includes Social Architects' political affiliation as an explanatory variable, adding the elasticity of Social

[17] 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 5: Welfare Weights and Preferences for Redistribution: Benchmarking

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	0.83
6	Republican + Elasticity of the weights	Yes	0.83

Notes: Each row of the table presents the root-mean-squared error (RMSE) of the predictions generated from a specification 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. A value of zero indicates that incomes should not be further redistributed. *Elasticity of the weights - std* is the elasticity of Social Architects' weights with respect to the income of the Recipients. textitRepublican 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.

Architects' welfare weights in the regression does not improve the predictive power of the specification. This result implies that the partisan gap in preferences for redistribution is likely driven by factors unrelated to welfare preferences. We explore the factors that explain the partisan gap in preferences for redistribution in Appendix Section D.4. We find that Social Architects' welfare weights explain 8% of the partisan gap in preferences for redistribution. The partisan gap is largely driven by concerns about externalities from inequality and concerns about the efficiency costs of taxation.

4.6.2 Which Factors Predict Preferences for Redistribution?

We explore whether Social Architects' preferences for redistribution are driven by their welfare preference (captured by welfare weights) or by factors orthogonal to welfare preferences. We regress Social Architects' preferences for redistribution (explained in the previous section) on Social Architects' standardized elasticity of welfare weights and other variables, 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 6a presents the coefficient estimates of the regression. We find that Social Architects' elasticity of the weights predicts their preferences for redistribution. However, Social Architects' misperceptions play a role too. Social Architects who overestimate the level of taxes have less progressive preferences 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 preferences 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 preferences for redistribution. Social Architects' views about taxes and government also play a role. Social Architects who think that higher taxes on high-income individuals hurt the economy have less progressive preferences for redistribution. Social Architects who believe in trickle-down economics have more progressive preferences 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 preferences for redistribution.

Overall, we find that while Social Architects' welfare weights predict their preferences for redistribution, their misperceptions and views about taxes and government play a role too. We have the following result.

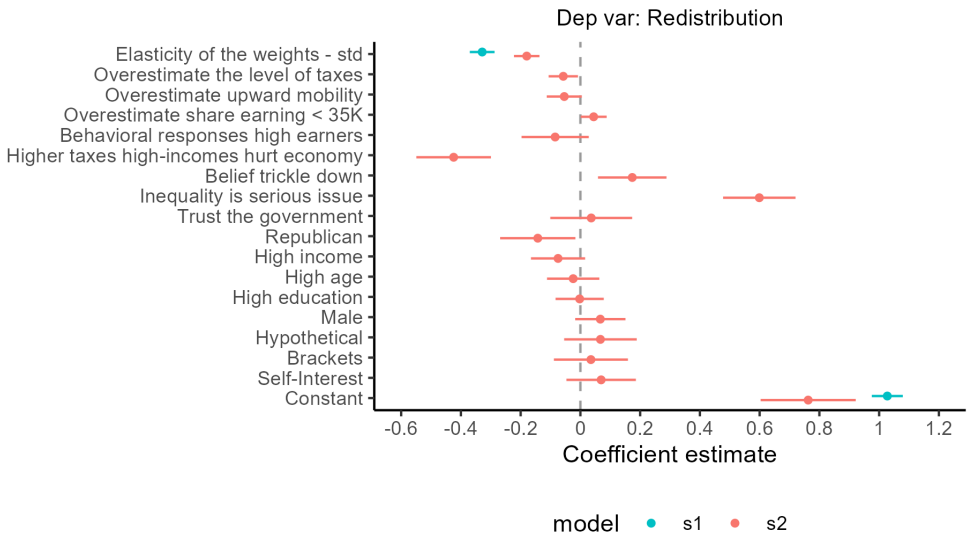
Result 5. *Social Architects' welfare weights, misperceptions, and views about taxation and government jointly predict their preferences for redistribution.*

4.6.3 Do Welfare Weights Capture Factors Orthogonal to Welfare Preferences?

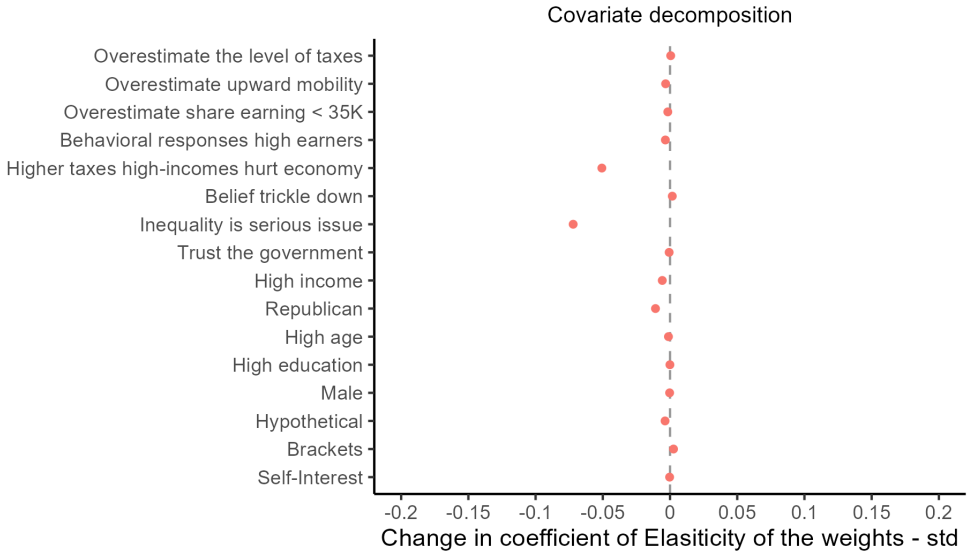
In the previous sections, we showed that Social Architects' welfare weights predict their preferences for redistribution. How much of this correlation is driven by factors orthogonal to welfare preferences? Factors that affect Social Architects' preferences for redistribution via their welfare weights are captured by our measure of welfare weights. Thus, the above question is akin to asking if the elicited welfare weights capture factors orthogonal to welfare preferences.

To answer this question, we use the covariate decomposition proposed by Gelbach (2016). First, we explore the overall variation in preferences for redistribution explained by welfare weights. To do so, we regress Social Architects' preferences for redistribution on their standardized elasticity of the weights. The results are presented in Model s1 in Figure 6a. We find that the coefficient estimate of the elasticity of the weights is -0.32.

Next, we explore the change in the coefficient estimate of the elasticity of the weights when we add variables (capturing other factors) to the regression. Details about the other factors can be found in Appendix Section A. The results are presented in Model s2 in Figure 6a. 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



(a) Coefficient Estimates



(b) Covariate Decomposition

Figure 6: Predictors of Preferences 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 coefficient 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.

halves in Model s2 when we account for other factors in the regression, indicating that a part of the variation in preferences for redistribution explained by welfare weights is

driven by these other factors.

How much of the change in the coefficient estimate of the elasticity of the weights (between Model s1 and s2) is due to each of the other variables? Figure 6b 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 adding this variable to the regression in Model s1 leads to a reduction, in absolute value, in the coefficient estimate of the elasticity of the weights. 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' preferences 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 result.

Result 6. *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' welfare weights do not capture many variables that are captured by their preferences for redistribution. This result suggests that using the experimental measure of welfare weights can be useful in settings where we want variation in welfare preferences rather than in factors orthogonal to welfare preferences.

4.7 Welfare Weights and Confidence

At the end of six decision screens, Social Architects in Wave 2 were asked how confident they were that the choices they made in the previous screens reflected what they really thought. We hypothesized that Social Architects who made mistakes in the experiment or made choices inconsistent with their preferences (possibly due to a poor understanding of the task) would report lower confidence levels. Responses to this question could range from "1: Not confident at all" to "5: Completely confident." Our data shows that 83% of the Social Architects reported a confidence level of 4 or 5, indicating that the majority of Social Architects are confident that their choices accurately reflect what they really think.

Next, we explore whether Social Architects' confidence correlates with their welfare weights. We regress the log of the welfare weights assigned by Social Architects on the log of the incomes of the Recipients, *High Confidence*, and the interaction term of the log of the incomes of the Recipients with *High Confidence*. *High Confidence* is an indicator

variable taking a value of 1 if a Social Architect's confidence is above the median and 0 otherwise. The regression is weighted using sampling weights. We find that Social Architects with high confidence have less progressive welfare weights relative to those with low confidence.[18] However, the effect is not statistically significant.

Overall, we find that Social Architects have a high level of confidence that their decisions reflect their preferences and that Social Architects' confidence is not correlated with their welfare weights. Our results help validate that our experimental design is largely capturing people's preferences rather than noise.

5 Comparing Welfare Weights

In Section 5.1, we describe the range of the general population that is used for comparisons. In Section 5.2, we compare the welfare weights of the general population to the weights implied by the income tax schedule and transfer policies in the U.S. In Section 5.3, we compare the welfare weights of the general population to the weights commonly used in the optimal policy literature. Finally, in Section 5.4, we explore the implications of the different sets of welfare weight for the optimal labor income taxes in the U.S.

5.1 Range of General Population Weights for Comparisons

We identify the range of ν across treatments that characterize the average general population weights. This range is computed based on Treatments Loss $\times$ 70K, Gain $\times$ 70K, and Base. We exclude the remaining treatments for various reasons.

Treatments Loss $\times$ 500K and Gain $\times$ 500K are excluded because these treatments have a limited range of possible values of ν. This limitation is discussed in Section 4.3. We exclude Treatments Brackets and Self-Interest because these treatments were implemented to test whether Social Architects' self-interest motives play a role. We follow the tradition in welfare economics of exploring people's views unconfounded by self-interest motives. Finally, we exclude Treatment Hypothetical because we conjecture that Social Architects in this treatment may not have considered the trade-offs carefully. We included Treatment Hypothetical to test whether the welfare weights elicitation can potentially be done with hypothetical decisions.

The range of ν is computed by estimating the regression specified in Equation (6) separately in each treatment. These regressions are weighted using sampling weights. Column (2) in Table A4 in Appendix Section C.2 presents the estimated values of ν. The

[18] ν is 0.07 higher for those with high confidence relative to those with low confidence.

range of ν based on our three preferred treatments is given by $\nu \in (-0.37, -0.53)$.

5.2 Comparing to Weights Implied by Tax and Transfer Policies

We compare the welfare weights of the general population to the weights implied by the income tax schedule in the U.S. and to the weights implied by transfer policies in the U.S. The welfare weights derived from the tax schedule and transfer policies reflect politicians' aggregation of societal welfare weights, potentially influenced by political economy considerations.

The weights implied by the income tax schedule, which are estimated by Hendren (2020), are derived based on the optimal income tax formula. This 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, which calculates the inverse-optimum welfare weights as a function of marginal tax rates and other relevant objects. The inverse-optimum welfare weights that make the current tax schedule optimal can be obtained by substituting the actual marginal tax rates into the inverse-optimum tax formula. Hendren (2020) estimates the inverse-optimum welfare weights implied by the income tax schedule using the universe of tax returns in 2012. Individuals' tax liabilities are 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 from the framework outlined in Hendren and Sprung-Keyser (2020). 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, 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 the tax schedule if and only if $s^* g(z^*) \geq c$. Rewriting this expression yields the following equation:

$$MVPF = \frac{s^*}{c} \geq \frac{1}{g(z^*)}. \tag{9}$$

Therefore, the inverse optimum weight implied by a transfer policy is obtained by taking the inverse of the policy's Marginal Value of Public Funds (MVPF). Hendren and Sprung-Keyser (2020) provide estimates of the MVPF of various policies. We use their estimates to compute the weights implied by these policies. From the set of policies studied

by Hendren and Sprung-Keyser (2020), we focus on taxes, cash transfers, and in-kind transfers, as they are the policies most similar to those in our study. Additionally, we restrict the sample to policies with a positive MVPF, as our theoretical framework and experimental design cannot accommodate negative welfare weights. The selected policies are listed in Table A6 in Appendix Section C.2.

The sets of welfare weights we are comparing correspond to different time periods. Furthermore, the welfare weights implied by the income tax schedule are based on pre-tax incomes while the welfare weights estimated in our experiment are based on disposable incomes. 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 the incomes of the Recipients.

We find that the elasticity of the weights implied by the income tax schedule is -0.10. The average weights of the general population (elasticity estimates ranging from -0.37 to -0.53) are about 3.7 to 5.3 times more progressive than the weights implied by the income tax schedule. The elasticity of the weights implied by transfer policies is -0.27, which is much closer to the estimates of the general population. The average weights of the general population are about 1.4 to 2 times more progressive than the weights implied by transfer policies. This leads to the following result.

Result 7. *The average 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 U.S.*

Figure 7 plots the interpolated welfare weights against the disposable income distribution. From percentiles 1 to 30, the general population weights are higher than the weights implied by the income tax schedule and transfer policies. Around the 30^{th} percentile, which corresponds to an average disposable income of \$35,697, the three sets of weights start to converge. At higher percentiles, the general population weights are lower than the weights implied by the income tax schedule and transfer policies.

Decomposing the Gap

What explains why the welfare weights of the general population are more progressive than the weights implied by the income tax schedule? To answer this question, we conduct a series of checks. First, we explore alternative assumptions about the aggregation of societal welfare weights. Under the median voter theorem, politicians' aggregation of welfare weights corresponds to the median welfare weights. We test whether the gap can

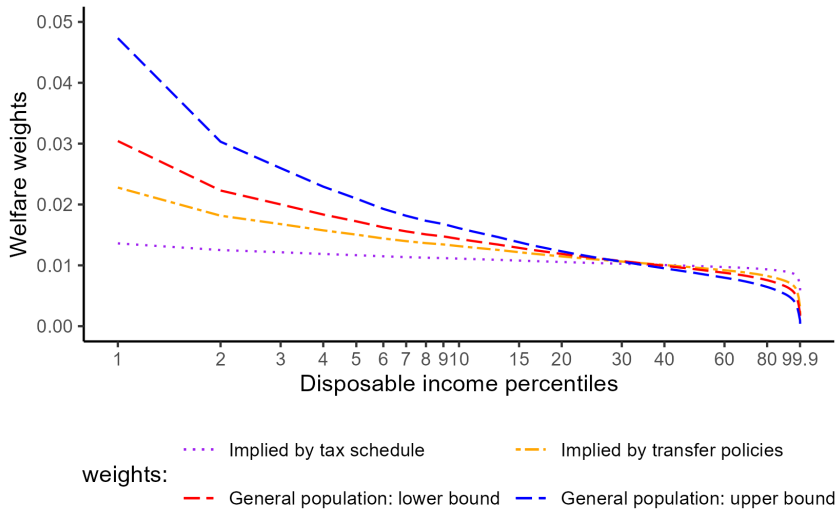


Figure 7: Average General Population Weights and Inverse-optimum Weights

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 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 analysis can be found in Appendix Section E.2.

be closed by comparing the weights implied by the tax schedule to the median welfare weights of the general population. We compute the weighted median value of ν using Equation (8), weighting the sample using sampling weights. We find that the estimated value of ν is -0.45, suggesting that the general population welfare weights remain more progressive than the welfare weights implied by the tax schedule.

Second, we explore alternative assumptions about the role of self-interest motives. We hypothesized that politicians may be aggregating the preferences of individuals who are guided, at least partly, by self-interest motives. However, our treatments were designed to minimize the influence of self-interest motives. Would allowing for self-interest motives to play a role close the gap between the average general population welfare weights and the welfare weights implied by the income tax schedule? In Section 4.3, we find that Social Architects' welfare weights are more progressive in Treatment Self-Interest relative to Treatment Brackets. This suggests that allowing self-interest motives to play a role would widen the gap.

Third, we consider alternative assumptions about the group of individuals that the politicians consider when aggregating welfare weights. In particular, politicians may be considering only the voting population, not the entire population. We test whether the

gap can be closed by comparing the welfare weights implied by the tax schedule to the weighted median welfare weights, weighting the sample by sampling weights, as well as the likelihood of being a registered voter.[19] We find that the weighted median value of ν is -0.47, once again suggesting that the general population welfare weights are more progressive than those implied by the tax schedule.

Finally, we explore the hypothesis of "elite capture," namely that politicians overweight the interests of high-income individuals when aggregating societal welfare weights. That is, they assign higher "aggregation weights" to high-income individuals. To test this hypothesis, we compute the median welfare weights, weighting the sample by sampling weights and 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 median estimate 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, there is some evidence that the gap between the general population weights and the weights implied by the income tax schedule is partly driven by politicians overweighting the interests of higher-income individuals. This explanation is consistent with the evidence in the literature suggesting that implemented policies are often more likely to reflect the interests of higher-income individuals relative to lower-income individuals (e.g., Gilens, 2005; Gilens and Page, 2014).

Overall, we find that the general population weights differ from the weights implied by the tax schedule partly because the latter places a higher weight on high-income individuals when aggregating societal welfare weights. However, this explanation does not close the gap. This leads to the following result.

Result 8. *The general population welfare weights differ from the welfare weights implied by the tax schedule partly because the latter places a higher weight on the welfare weights of high-income individuals when aggregating societal welfare weights.*

5.3 Comparing to Weights Used in the Literature

We compare the average welfare weights of the general population to the welfare weights frequently used in the optimal policy literature, which are inversely proportional to the disposal incomes of the Recipients (y^ν, with $\nu = -1$). This corresponds to utilitarian welfare weights assuming log utilities, which we refer to as log-utilitarian weights.

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

We find that the log-utilitarian welfare weights are 1.8 to 3.3 times more progressive than the welfare weights of the general population ($\nu \in (-0.37, -0.53)$). Figure A7 in Appendix Section C.2 plots the general population welfare weights and the log-utilitarian weights. We have the following result.

Result 9. *The welfare weights commonly used in the literature, which correspond to utilitarian welfare weights assuming logarithmic utilities, are about 1.8 to 3.3 times more progressive than the average welfare weights of the general population.*

5.4 Calibrating Optimal Labor Income Taxes

We explore the implications of the different sets of welfare weights for the optimal labor income taxes in the U.S. We use the tax formula derived in Saez and Stantcheva (2016) and Saez (2001), which is given as

$$T'(z) = \frac{1 - \bar{G}(z)}{1 - \bar{G}(z) + e(z) \cdot \alpha(z)} \tag{10}$$

where $T'(z)$ is the marginal tax rate for income z. $e(z)$ is the average 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). The optimal marginal tax rate is decreasing in the elasticity of earnings to reduce the extent to which taxes distort people's labor supply. The local Pareto parameter $\alpha(z)$ is given by $\alpha(z) = zh(z)/[1 - H(z)]$, where $h(z)$ is the income density and $H(z)$ is the cumulative distribution of income. The local Pareto parameter measures the shape of the income distribution. The optimal marginal tax rate is decreasing in $\alpha(z)$. Intuitively, having a higher marginal tax rate at an income z with a thin density minimizes the distortions to those earning z while leaving unaffected the incentives of those earnings above z. Finally, $\bar{G}(z)$ is the average welfare weight assigned to individuals index by j earning above z. The formula is given by

$$\bar{G}(z) = \frac{\int_{j:z_j \geq z} g_j \, d_j}{Prob(z_j \geq z) \cdot \int_j g_j \, dj}. \tag{11}$$

To obtain $\bar{G}(z)$, we sum the welfare weights for all incomes above (including) z and divide the sum by the share of individuals earning above (including) z. In the formula, the sum of the weights across all individuals given by $\int_i g_i \, di$ equals one. The optimal marginal tax rate is increasing in $\bar{G}(z)$. The higher the welfare weights assigned to Recipients earning above z, the higher the marginal tax rate for the individuals earning z.

We calibrate the sufficient statistics in the tax formula as follows. We set $\alpha(z) = 1.5$ for simplicity. We set $e(z) = 0.25$, which is a mid-range estimate for the elasticity of taxable income (Saez et al., 2012). Next, we calibrate $\bar{G}(z)$. We interpolate the average welfare weights using the function $\bar{g}(z) = (z - T(z))^{\nu}$, where ν is a parameter that governs the progressivity of the weights, and $(z - T(z))$ is the after-tax income of the individuals. We re-normalize the weights such that they sum to 1 and plug them in Equation (11). Different estimates of welfare weights can be obtained by plugging in different values for ν.

Figure 8 plots the optimal marginal tax rates for different estimates of welfare weights, along with the actual marginal tax rates based on the 2019 tax schedule for single filers. The x-axis represents the percentiles of the labor income distribution, which has been smoothed to reduce noise. Details on the construction of the income distribution can be found in Appendix Section E.4.

We find that the slope of the optimal marginal tax schedule based on the welfare weights implied by the tax schedule is very similar to the actual marginal tax schedule. This helps us validate the exercise of obtaining optimal tax schedules.

The optimal marginal tax rates based on log-utilitarian welfare weights rise steeply and then plateaus around the 5^{th} percentile of the income distribution. We find that the average optimal marginal tax rates obtained by plugging in the upper bound of the general population weights ($\nu = -0.53$) are about 13 percentage points lower than the average optimal marginal tax rates obtained by plugging in utilitarian welfare weights typically used in the literature ($\nu = -1$). The difference is 22 percentage points if we instead use the lower bound of the general population weights ($\nu = -0.37$).

The average optimal marginal tax rates based on the weights implied by transfer policies ($\nu = -0.26$) is about 7-16 percentage points lower than the average optimal marginal tax rates based on the general population weights. The optimal marginal tax rates implied by the income tax schedule ($\nu = -0.1$) or the actual income tax schedule in the U.S. is 26-35 percentage points lower than the average optimal marginal tax rates based on the general population weights.

Overall, our results suggest that the optimal marginal tax rates are much lower when calibrated with the general population weights relative to the log-utilitarian weights frequently used in the literature. The optimal marginal tax rates implied by the general population weights are better approximated by the optimal marginal tax rates implied by the log-utilitarian welfare weights than the weights implied by the income tax schedule in the U.S. The optimal marginal tax rates implied by the general population weights are similar to those implied by transfer policies in the U.S. We have the following result.

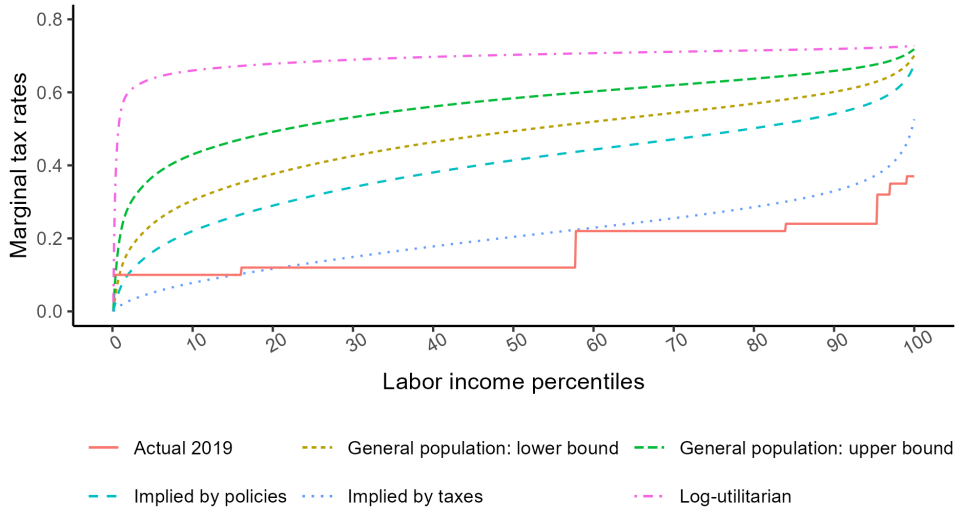


Figure 8: Marginal Tax Rates

Notes: The figure plots marginal tax rates against percentiles of the labor income distribution. The figure plots the optimal marginal tax rates obtained using the inverse-optimum weights implied by the income tax schedule computed by Hendren (2020) ($\nu = -0.1$), inverse-optimum weights implied by transfer computed by Hendren and Sprung-Keyser (2020) ($\nu = -0.26$), lower-bound of the general population weights ($\nu = -0.37$), upper-bound of the general population weights ($\nu = -0.53$), and log-utilitarian weights ($\nu = -1$). The figure also plots the actual marginal tax rate in 2019 for single filers. The labor income distribution is obtained from Piketty et al. (2018) (*plinc*). We restrict the sample to positive labor incomes.

Result 10. *The optimal marginal tax rates are much lower when calibrated with the general population weights relative to the log-utilitarian weights frequently used in the literature. The optimal marginal tax rates implied by the general population weights are better approximated by the optimal marginal tax rates implied by the log-utilitarian welfare weights than the weights implied by the income tax schedule in the U.S. The optimal marginal tax rates implied by the general population weights are similar to those implied by transfer policies in the U.S.*

6 Welfare Weights and Ideals

6.1 Welfare Weights and Non-Welfarist Motives

In the traditional “welfarist" approaches, welfare weights can depend on characteristics directly affecting Recipients’ utility functions (e.g., Recipients’ disability status). On the other hand, in “non-welfarist" approaches, welfare weights can depend on characteristics that do not directly affect Recipients’ utility functions (e.g., Recipients’ parental income). In this section, we explore whether Social Architects’ welfare weights are driven by non-welfarist motives. This is an important exercise because the value of the approach developed by Saez and Stantcheva (2016) lies in incorporating non-welfarist motives in the welfare weights. Accounting for non-welfarist motives allows the welfare weights to

capture views such as equality of opportunity, accounting for horizontal-equity concerns, or accounting for freeloaders.

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 not deserve their current income and do not need their current income," (ii) "deserve their current income but do not need their current income," (iii) "do not 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 motives in which welfare weights depend on the needs of the Recipients. The word "deserve" captures all non-welfarist motives.

We present the results in Appendix Section D.5. Our results suggest that the Social Architects' welfare weights are guided jointly by non-welfarist and welfarist motives. Combining this finding with the finding that the Social Architects' weights are heterogeneous (documented in Section 4.4) supports our general sufficient-statistics approach that can incorporate various welfarist and non-welfarist ideals.

Result 11. *Social Architects' welfare weights are guided by welfarist and non-welfarist motives.*

6.2 Welfare Weights and Source of Income

Welfare weights are general enough to capture various ideals. In this section, we test whether Social Architects' welfare weights capture one important ideal document in the literature: redistribution based on the source of income. Several studies have documented that people are willing to accept inequalities if the individuals have earned their income through effort, but not if they have earned their income through luck (e.g., **?**). To capture this ideal, we elicited Social Architects' beliefs about whether high-income individuals are rich due to luck or effort.

We regress the log of the welfare weights assigned by Social Architects on the log of the incomes of the Recipients, *Person Rich due to effort*, and the interaction term of the log of the incomes of the Recipients with *Person Rich due to effort*. *Person rich due to effort* is an indicator variable taking a value of 1 if a Social Architect indicates that a person is rich due to effort and 0 if the Social Architect indicates that a person is rich due to luck. We find that Social Architects who think that higher-income individuals accrued their income due to effort have less progressive welfare weights relative to those who think that higher-income individuals accrued their income due to luck. This suggests that Social Architects' beliefs

about the source of income influence their welfare weights.

7 Discussion

Evaluating the desirability of a reform involves weighing the gains of the winners against the losses of the losers using welfare weights. Welfare weights measure the value society places on providing an additional dollar of consumption to individuals. These weights are general enough to capture various ideals, such as equality of opportunity or poverty alleviation. In our project, we elicit the welfare weights assigned by the general population of the U.S.

We conducted two real-stakes online experiments ($N \approx 4000$) with samples of the general population in the U.S. In the experiments, participants in the role of "Social Architects" make choices between various reforms that are used to elicit the welfare weights assigned to seven real "Recipients." We adopt a sufficient statistics approach in this paper. A Social Architect's assessment of the welfare implications of a reform is a function of "high-level" estimable statistics. This approach does not require specifying and uncovering the underlying ideals that guide the Social Architects.

We find that the average welfare weights of the general population are progressive, indicating a desire for additional redistribution beyond that achieved by the current tax and transfer system. We find that the general population welfare weights are more progressive than the weights implied by the income tax schedule in the U.S., but similar to the weights implied by transfer policies in the U.S. Finally, we find that the general population weights are less progressive than the weights frequently used in the optimal policy literature.

7.1 Aggregation of Welfare Weights

A key question in our setting relates to the aggregation of Social Architects' welfare weights. In our paper, we aggregate the welfare weights of the Social Architects by computing the average or the median of the welfare weights. There are two potential concerns regarding aggregation.

First, computing the simple average of the welfare weights of Social Architects may not be the normatively correct aggregation. Exploring different ways to aggregate the welfare weights 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.

Second, it is unclear whether the aggregate welfare weights of the Social Architects should be used in policy formulas or if the individual welfare weights of Social Architects should be plugged into policy formulas and the resulting policies should be aggregated

across Social Architects. One advantage of aggregating the welfare weights of the Social Architects is that the average welfare weights may not be affected by measurement error or Social Architects' misperceptions if the measurement error and Social Architects' misperceptions have a sum of zero.

7.2 Limitations of our Approach

Our sufficient-statistics approach to eliciting welfare weights and the "small-reform" approach to taxation used by Saez and Stantcheva (2016) has several limitations. First, the welfare weights estimated in our paper cannot be used to evaluate non-marginal ("large") reforms.[20] 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.

Second, our approach is unsuitable in settings involving general equilibrium effects. In the presence of general equilibrium effects, the effects of a reform might extend beyond those who were the direct target of the reform. This makes the assessment of the desirability of the reform challenging. However, this is a general limitation of the sufficient-statistics approach and also extends to other objects in the tax formulas, such as the elasticity of taxable income.

[20] 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).

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.

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* .
URL: *https://doi.org/10.1093/qje/qjad007*

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

Charité, J., Fisman, R. and Kuziemko, I. (2015), 'Reference Points and Redistributive Preferences: Experimental Evidence', *National Bureau of Economic Research* .

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* .

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. (2022), Confidence, self-selection and bias in the aggregate, Technical report, National Bureau of Economic Research.

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. (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.

Fong, C. M. and Poutvaara, P. (2019), 'Redistributive politics with target-specific beliefs', *Available at SSRN 3360749* .

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

Gilens, M. (2005), 'Inequality and democratic responsiveness', *Public Opinion Quarterly* **69**(5), 778–796.

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. (2021), 'Designing information provision experiments', *Journal of Economic Literature* .

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.

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.

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

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.

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

Schokkaert, E. and Devooght, K. (2003), 'Responsibility-sensitive fair compensation in different cultures', *Social Choice and Welfare* **21**(2), 207–242.

Singhal, M. (2021), 'Quantifying preferences for redistribution', *Working paper* .

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.

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. (2017), 'Popular acceptance of inequality due to innate brute luck and support for classical benefit-based taxation', *Journal of Public Economics* **155**, 54–63.

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 **48**

B Pre-registration **50**

C Additional Tables and Figures **52**
C.1 Additional Figures 52
C.2 Additional Tables 56

D Additional Analysis **62**
D.1 Role of Social Architect's Own Income 62
D.2 Treatment Effects 63
D.3 Welfare Weights and Preferences for Redistribution 64
D.4 Partisan Gap in Preferences for Redistribution 66
D.5 Social Architects' Motives 67

E Details on Analyses **70**
E.1 Details on the Figure Plotting the Incomes of the Recipients 70
E.2 Details on the Interpolation of Welfare Weights 70
E.3 Construction of Likelihood of Registered Voters 71
E.4 Calibration of the Tax Formula 71

F Proofs **72**
F.1 Incentive Compatibility of the Staircase Procedure 72

G Instructions - Wave 1 **75**

H Instructions - Wave 2 **89**

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 policy views on 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 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 misperceptions variables and then standardize 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 second-poorest Recipient 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 variable such that it has a mean of 0 and a standard deviation of 1.

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 variable such that it has a mean of 0 and a standard deviation of 1.

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 question captures a unique mechanism that may help explain people's policy preferences. We draw these questions from Stantcheva (2020). The prompt for the question on inequality being a serious issue was 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 their 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 in government: We elicit Social Architects' level of trust in the U.S. government. In particular, we ask them 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.[21] We report three deviations from the pre-registration in the data collection and sample restrictions. 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 dropped participants with multiple survey responses (6 participants in Wave 1 and 7 participants in Wave 2). We did not pre-register this sample restriction as we did not expect participants to take the survey multiple times. Next, 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 had initially pre-registered using the log of the income and the log of the income squared as controls before Wave 1. We had 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’ elasticity of the welfare weights with respect to the incomes of the Recipients using OLS regressions (by minimizing the sum of squared errors). We had pre-registered estimating the elasticity of the welfare weights 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. While 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 easily transferable to other settings.

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

5. In the section (Appendix Section D.1) exploring the role of Social Architects' own income on their assigned weights, we present regressions in which the explanatory 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 how the role of self-interest motives differs between Social Architects with different incomes.

6. We explore the role of self-interest motives (Table A10 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. In our applications, we exclude Treatments Loss $\times$ 500K and Gain $\times$ 500K. We had initially pre-registered using these treatments since we were not aware of the extent to which participants' choices in these treatments would lead to biased estimates of welfare weights because of the bounds in the welfare weights. We did not pre-register which of the treatments in Wave 2 we would use for the applications.

9. The results in Table A7 were not pre-registered. Some regressions in Section 4.6.1 and Section D.3 were not pre-registered.

C Additional Tables and Figures

C.1 Additional Figures

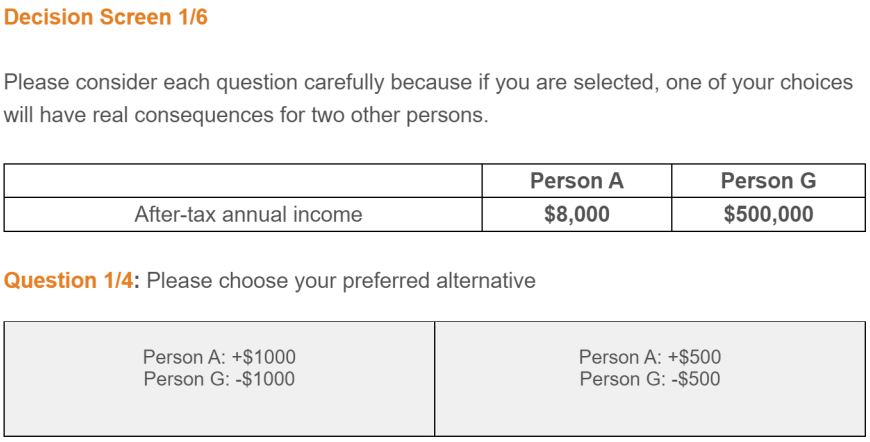


Figure A1: Screenshot of a Question Asked to Social Architects

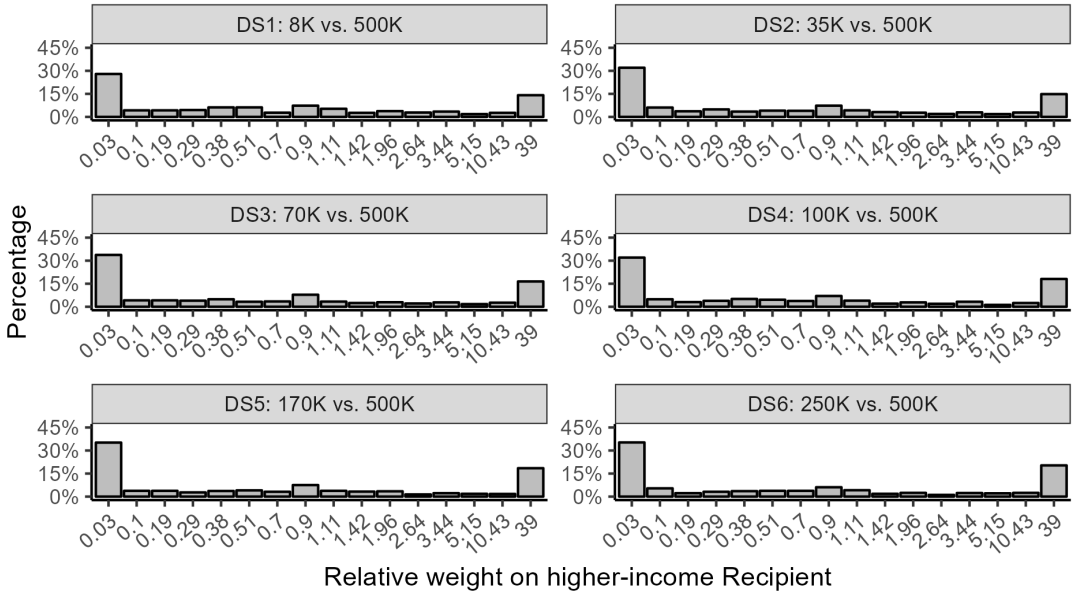


Figure A2: Distribution of $\tilde{g}$ Across Decision Screens

Notes: The figure presents the distribution of the welfare weight assigned to the higher-income Recipient relative to the lower-income Recipient across the six decision screen ("DS"). The relative weight $\tilde{g}$ is calculated using Equation (5). Social Architects are presented with a different pair of Recipients in each decision screen. For example, in the first decision screen, Social Architects face Recipients earning $8,000 and $500,000. For half the participants, the order of the decision screens was reversed. The sample is restricted to treatments where the Recipient common across decision screens has an income of $500,000.

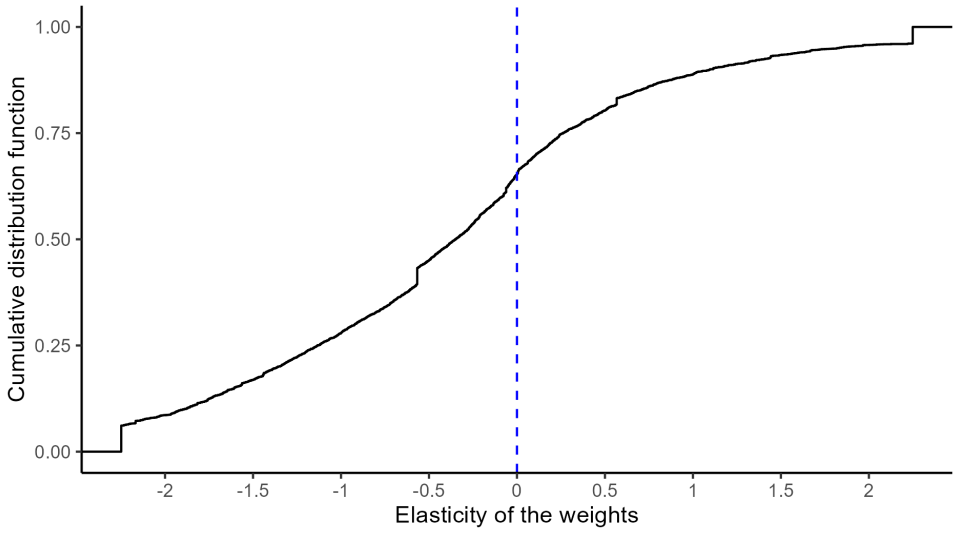


Figure A3: Distribution of the Elasticity of Social Architects' Welfare Weights

Notes: The figure presents the cumulative distribution function (CDF) of the parameter estimate $\hat{v}$ obtained from the following regression $log(g(R_j)) = \beta_0 + vlog(\textit{recipient income}_j) + \epsilon_j$, where $g(R_j)$ is the weight assigned by a given Social Architect to Recipient j, and $\textit{recipient income}_j$ is the income of Recipient j.

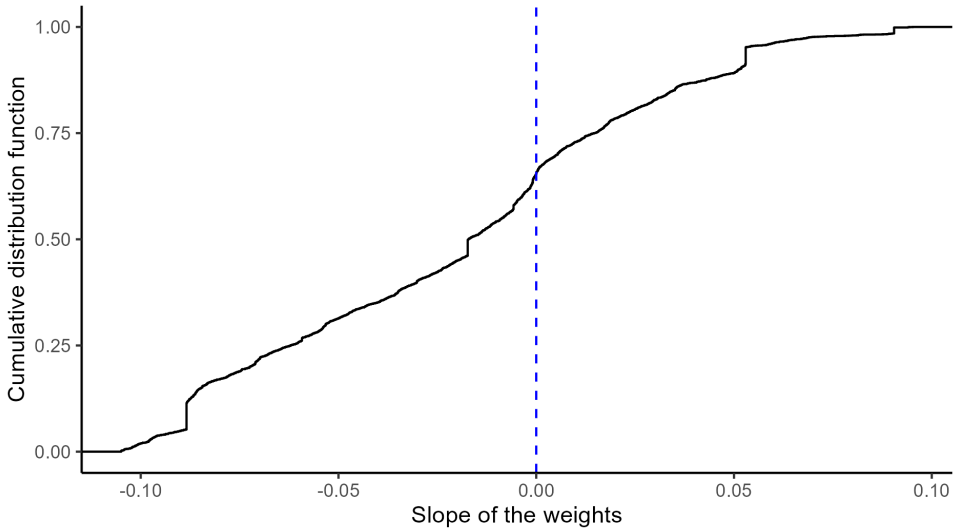


Figure A4: Distribution of the Slope of Social Architects' Welfare Weights

Notes: The figure presents the cumulative distribution function (CDF) of the parameter estimate $\hat{\beta_1}$ obtained from the following 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.

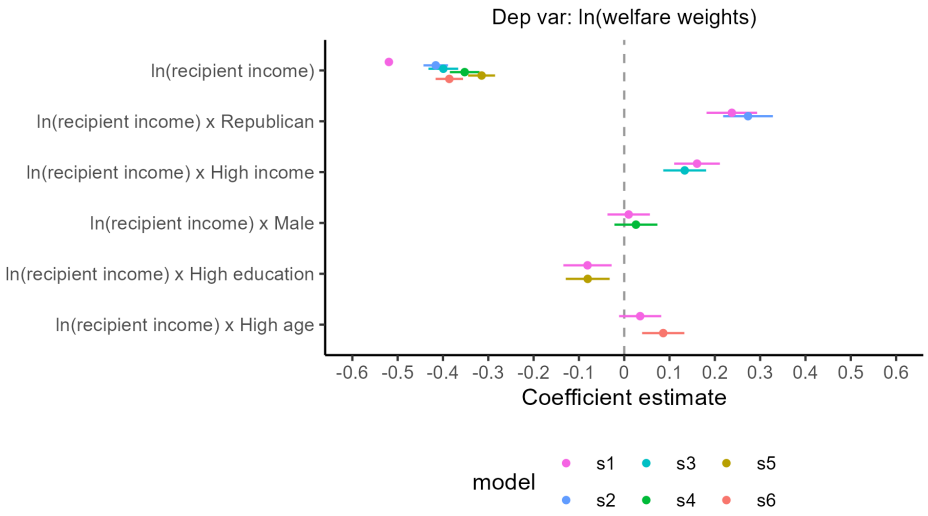


Figure A5: Social Architects' Welfare Weights and Characteristics

Notes: The figure presents coefficient estimates. The dependent variable is the log of the welfare weights assigned by Social Architects. In Model s1, 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. In Models s2 through s6, we include only one characteristic in each model and exclude treatment dummies. With the exception of the main effect of the log of the incomes of the Recipients, we do not present the main effects of the other variables. In Model s1, 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.

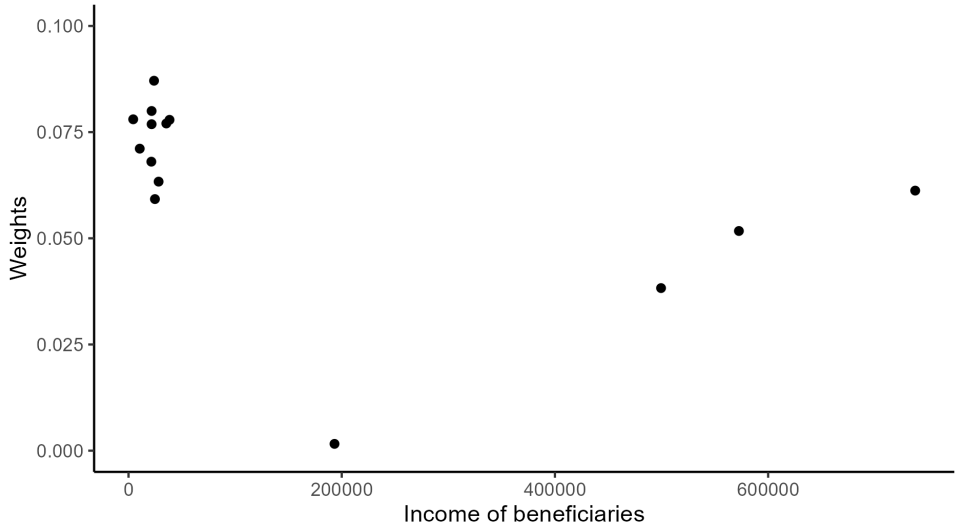


Figure A6: Weights Implied by Transfer Policies and Incomes of Beneficiaries

Notes: The figure plots the welfare weights implied by transfer policies against the average incomes of the beneficiaries of the policies. The weights implied by policies are computed using the data obtained from Hendren and Sprung-Keyser (2020) and are normalized to sum to 1.

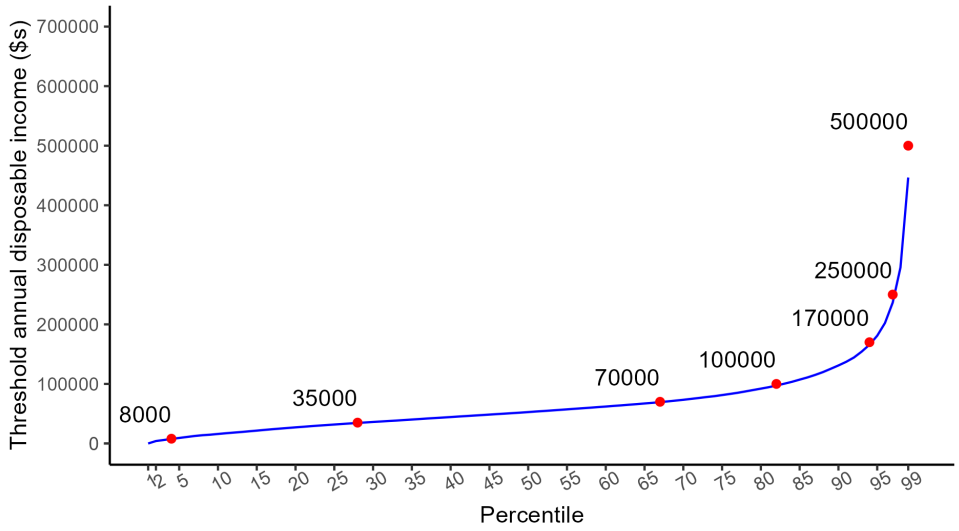


Figure A7: Average General Population Weights and Utilitarian Weights

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 log-utilitarian 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 income can be found in Appendix Section E.2.

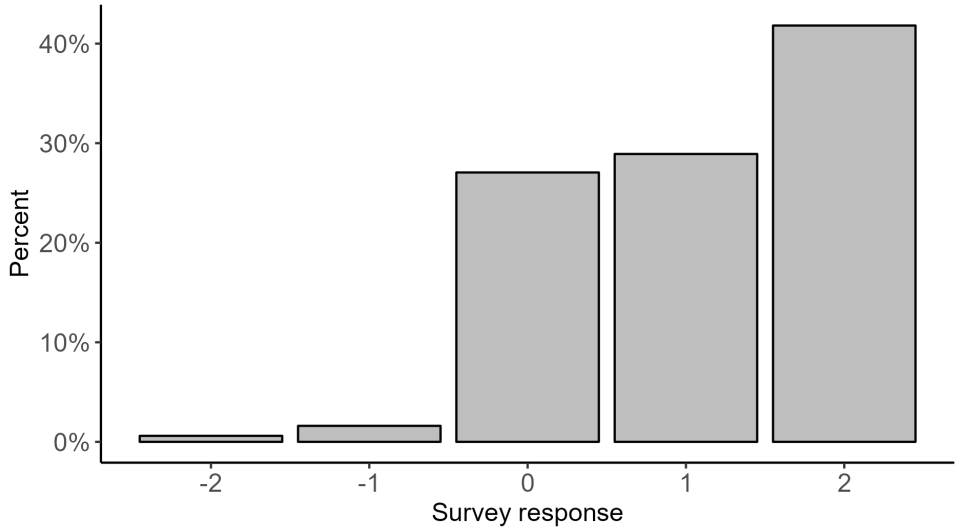


Figure A8: Frequency of Responses on the question *Redistribution*

Notes: The figure plots the frequency of responses on 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

	Base	Hypo	Brackets	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. The last column indicates the p-value corresponding to the F-statistic when testing if the treatment dummies are jointly significant.

Table A3: Patterns in the Welfare Weights

Variable	Full	Loss x 70K	Base	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 Base (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 Base. 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 following 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. The elasticity of the weights is the parameter estimate $\hat{v}$ obtained from the following 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 $\textit{recipient income}_j$ 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 A4: Elasticity of the Weights by Treatments

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

Notes: The table presents the mean and standard error of the estimated elasticity of Social Architects' welfare weights with respect to the incomes of the Recipients (ν). Each row presents regression estimates computed by estimating Equation (6) using the subset of Social Architects indicated in the column "Case." 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.

Table A5: Predictors of the Elasticity of the Weights

Dependent Variable:	Elasticity of the weights		
Model:	(1)	(2)	(3)
Constant	-0.505***	-0.676***	-0.683***
	(0.057)	(0.070)	(0.064)
Gain x 70K	-0.135*	-0.101	-0.125
	(0.079)	(0.086)	(0.079)
Loss x 500K	0.269***	0.303***	0.284***
	(0.061)	(0.069)	(0.062)
Gain x 500K	0.244***	0.259***	0.253***
	(0.061)	(0.068)	(0.061)
Base	-0.214***	-0.151*	-0.165**
	(0.075)	(0.084)	(0.076)
Hypothetical	-0.316***	-0.303***	-0.270***
	(0.079)	(0.084)	(0.078)
Brackets	0.208***	0.269***	0.262***
	(0.075)	(0.083)	(0.076)
Self-Interest	-0.010	0.023	0.043
	(0.074)	(0.082)	(0.075)
Screen order	0.281***	0.293***	0.290***
	(0.034)	(0.037)	(0.034)
Republican		0.242***	0.281***
		(0.044)	(0.042)
High income		0.174***	0.158***
		(0.039)	(0.037)
Male		0.009	0.017
		(0.037)	(0.035)
High education		-0.093**	-0.085**
		(0.042)	(0.040)
High age		0.035	0.038
		(0.036)	(0.035)
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 the incomes of the Recipients. *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 shown to 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 the policies chosen from Hendren and Sprung-Keyser (2020) along with their type and the average income of the Recipients of the policies.

D Additional Analysis

D.1 Role of Social Architect's Own Income

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

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

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 \$22,000. The other variables are similarly defined based on the following income brackets: (\$22000, \$53000], (\$53000, \$85000], (\$85000, \$135000], (\$135000, \$210000], (\$210000, \$375000], (\$375000, 10000000). These income brackets are the same as the ones used in Treatments Brackets and Self-Interest. To test if Social Architects assign a higher weight to Recipients with similar incomes than to other Recipients, we estimate a fixed-effects model with Social Architect fixed effects. In this regression, the identifying variation is across decision screens. We restrict the sample to the treatments excluding Treatment Self-Interest since this treatment incorporates self-interest motives by design. Here, we focus on understanding whether self-interest motives play a role in treatments where Social Architects are not affected by their decisions. We describe the role of self-interest in Treatment Self-Interest in Appendix Section D.2.

Figure A9 presents the coefficient estimates. Model s1 presents the specification described above. In Model s2, we present coefficient estimates in which 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 whose incomes are near Recipients 1, 2, and 4 assign a higher weight to Recipients with similar incomes relative to other Recipients. Social Architects with incomes near Recipient 1 assign a 10 percentage point higher weight to Recipient 1 relative to other Recipients. Interestingly, Social Architects with incomes near Recipient 3 (earning \$70000) 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 five, six, and seven. However, we are also underpowered to detect the effects for higher-income Social Architects since we have very few Social Architects with high incomes.

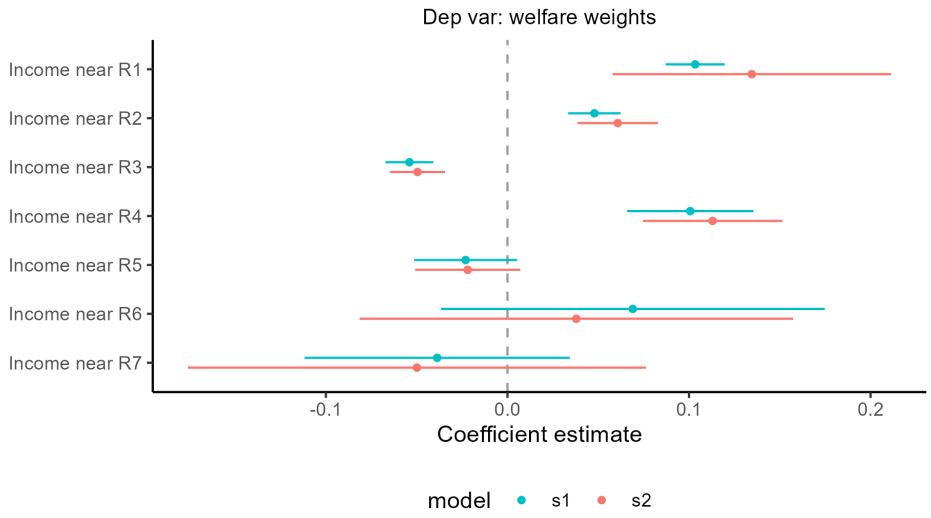


Figure A9: Weights Assigned to Recipients with Similar Incomes

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: ($22000, $53000], ($53000, $85000], ($85000, $135000], ($135000, $210000], ($210000, $37500], ($375000, ∞). In model s2, *Income near R1* takes a value of 1 if a Social Architect's income is plus or minus 0.2 times 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.

D.2 Treatment Effects

Comparing Treatments Base and Hypothetical

We explore one reason Social Architects may have more progressive welfare weight in Treatment Hypothetical relative to Treatment Base: 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 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 Base. 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 question, respectively. We test if Social Architects in Treatment Hypothetical choose one option throughout to finish 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 Base. Overall, we do not find evidence that Social Architects in Treatment

Hypothetical are less attentive in the survey.

Table A7: Comparing Treatments Base and Hypothetical

Dependent Variables:	Time spent (mins)	Always CR	Always VR
Model:	(1)	(2)	(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 question. *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 Base. The sample is restricted to Social Architects in Treatments Hypothetical and Base. 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

Exploring Self-Interest Motives

In the main text, we found that participants in Treatment Self-interest have more progressive weights than participants in Treatment Brackets, indicating that self-interest motives do play a role. In this section, we compare Social Architects' welfare weights in Treatment Brackets 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 A10 presents the results. Looking at the coefficient estimate of *ln(recipient income)* $\times$ *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 Bracket.

D.3 Welfare Weights and Preferences for Redistribution

We explore the empirical link between Social Architects' welfare weights and their preferences 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 preferences of redistribution administered in Wave 1. The first question asks Social Architects if they think the government should do more to reduce

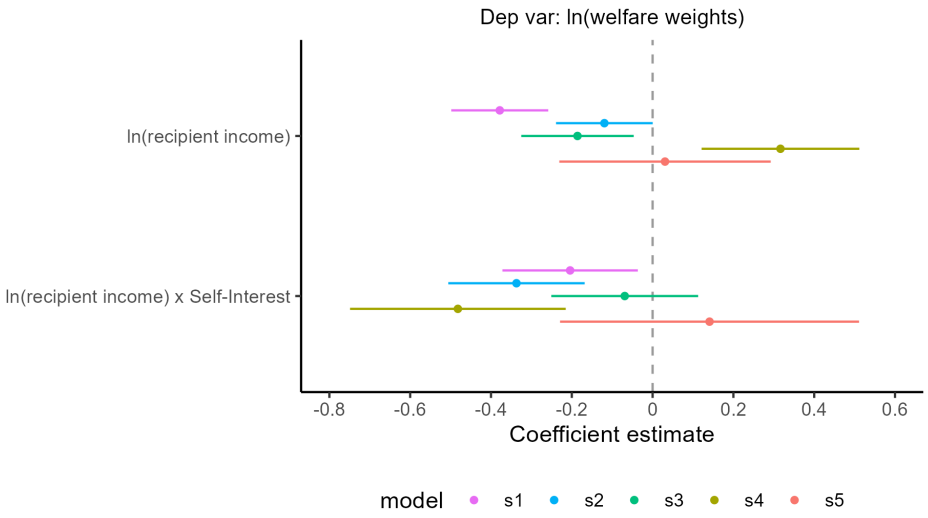


Figure A10: Social Architects' Weights and Self-Interest Motives

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 Brackets. The regressions use data from Treatment Brackets and Treatment Self-Interest. Models s1 to s5 include Social Architects from the following income brackets (0, \$22000], (\$22000, \$53000], (\$53000, \$85000], (\$85000, \$135000], (\$135000, ∞), respectively. We do not present the main effects of *Self-Interest*. Error bars are computed using HC3 standard errors.

income differences between the rich and poor. This question was used in the General Social Survey (GSS). We define the variable *Govt should do more*, which 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 reduce or increase the taxes on top-income earners. A modified version of this question was used by Cohn et al. (2019) and Kuziemko et al. (2015). We define the variable *Increase top-taxes*, which 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 against their stated political affiliation. We estimate a series of regressions in which the dependent variables are *Govt should do more* or *Increase top-taxes*. We regress Social Architects' preferences 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.[22] The lower the RMSE of a

[22]We divide the data into four sub-samples ($S(k), k \in 1,2,3,4$) with $k = 4$. For each sub-sample, we

specification, the higher the predictive power of that specification.

Table A8 presents the results. Looking at rows one, two, five, and six, we find that the elasticity of Social Architects' welfare weights is just as good a predictor of their preferences for redistribution as their stated political affiliation. We find similar results in rows three, four, seven, and eight, which include treatment dummies and background characteristics as controls in the regressions.

Table A8: Welfare Weights and Preferences for Redistribution: Benchmarking

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

Notes: Each row of the table presents the root-mean-squared error (RMSE) of a specification using a k-fold cross validation with $k = 4$. *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. *Increase top-taxes* takes values from 1 to 7, with higher values indicating a greater desire for the government to increase top-taxes. *Elasticity of the weights* is the elasticity of Social Architects' weights with respect to the income of the Recipients. *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.

D.4 Partisan Gap in Preferences for Redistribution

Our results in the main text suggest that Republicans have less progressive preferences for government redistribution relative to Democrats. In this section, we explore the factors that explain the partisan gap in preferences for redistribution using the covariate decomposition proposed by Gelbach (2016).

First, we explore the overall partisan gap in preferences for redistribution. We regress Social Architects' preferences for redistribution (variable *Redistribution*) on their political affiliation. The results are presented in Model s1 in Figure A11. The coefficient estimate

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.

of *Republican* in Model s1 is -0.69. This estimate reflects the overall partisan gap in preferences for redistribution.

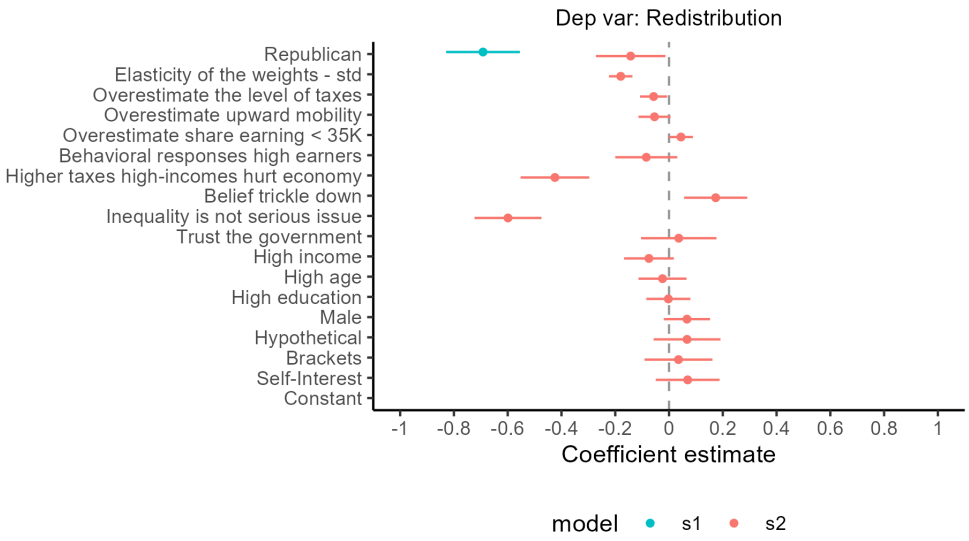
Next, we explore the change in the coefficient estimate of *Republican* when we add Social Architects' standardized elasticity of the weights and other variables (capturing other factors) to the regression. Details about the other factors can be found in Appendix Section A. The results are presented in Model s2 in Figure A11a. 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 that 79.4% of the partisan gap in policy preferences is driven by the other variables in the model.

How much of the change in the coefficient estimate of *Republican* (between Model s1 and s2) is due to each of the other factors? Figure A11b 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 adding this variable to the regression in Model s1 leads to a reduction, in absolute value, in the coefficient estimate of *Republican*. We find that about 8% of the partisan gap in preferences 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.

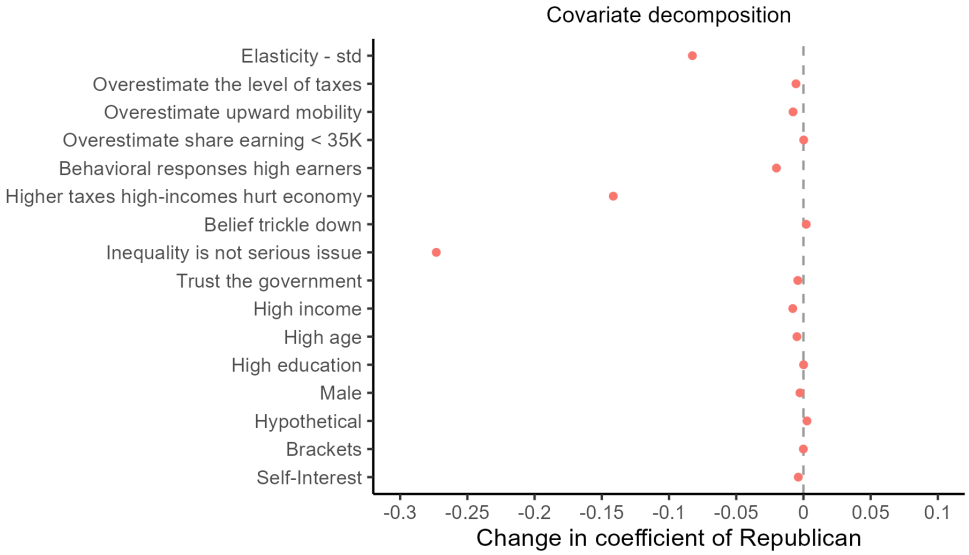
D.5 Social Architects' Motives

We explore how Social Architects' motives affect their welfare weights. We regress the log of the welfare weights assigned by Social Architects on 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.

Figure A12 presents the results. We find that the coefficient estimate of *ln(recipient income)* $\times$ *HI deserve but do not need* is greater than 0, implying that conditional on Social Architects believing that high-income individuals do not need their current income, Social Architects believing that high-income individuals deserve their current income have less progressive weights. We find a similar effect, qualitatively, of Social Architects' believing that high-income individuals deserve their income, conditional on believing that they need their current income. This can be seen by comparing the coefficient estimates of *ln(recipient income)* $\times$ *HI do not deserve but need* and *ln(recipient income)* $\times$ *HI deserve and need*. Looking



(a) Regression estimates



(b) Covariate Decomposition

Figure A11: Partisan Gap in Preferences 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. We reverse code *Inequality is serious issue* in these regressions. Models s1 and s2 are weighted using sampling weights. Error bars are computed using HC3 standard errors. Panel (b) presents the effect of each factor 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.

at the coefficient estimate of *ln(recipient income)* $\times$ *HI do not deserve but need*, we find that conditional on Social Architects believing that high-income individuals do not deserve

their current income, Social Architects believing that high-income individuals need their current income have less progressive weights. We find a similar effect, qualitatively, of Social Architects' believing that high-income individuals need their income, conditional on believing that they deserve their current income. Social Architects' beliefs about the needs and deservingness of low-income individuals play a smaller role in driving Social Architects' welfare weights.

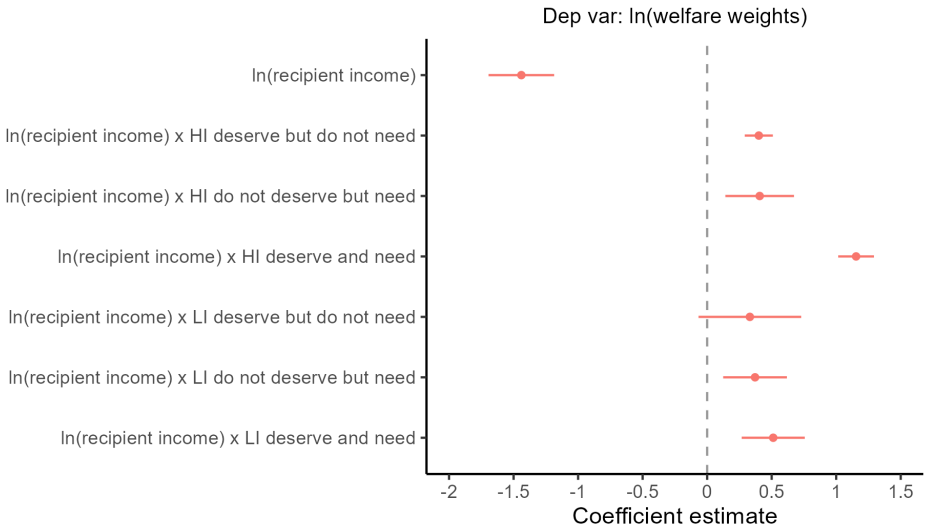


Figure A12: Social Architects' Weights and Motives

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. With the exception of the 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.

E Details on Analyses

E.1 Details on the Figure Plotting the Incomes of the Recipients

We present the details on the construction of the income distribution for the figure plotting the incomes of the Recipients. Data on the income distribution was obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). To simplify the analysis, we consider all individuals in the data as single filers, disregarding their tax filing status or whether they filed taxes at all. Our chosen measure of income is disposable income (*diinc*), which includes in-cash and in-kind transfers.

Step 1: We exclude individuals with negative disposable incomes.

Step 2: We partition the data into percentiles (1, ..., 99) based on the distribution of disposable incomes. Specifically, we determine the disposable income thresholds that group individuals into their respective percentiles.

Step 3: We assign each of the seven Recipients to the percentile that corresponds to their income.

Step 4: We plot these disposable income thresholds against the corresponding percentiles. In this plot, we also indicate the incomes corresponding to the Recipients.

E.2 Details on the Interpolation of Welfare Weights

We present the details on the analysis for the figures plotting the welfare weights against the income distribution. Data on the income distribution was obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). To simplify the analysis, we consider all individuals in the data as single filers, disregarding their tax filing status or whether they filed taxes at all. Our chosen measure of income is disposable income (*diinc*), which includes both cash-based redistribution and in-kind transfers.

Step 1: We exclude individuals with negative disposable incomes.

Step 2: We partition the data into percentiles (1, . . ., 99, 99.9, 99.99) based on the distribution of disposable incomes.

Step 3: We compute the average disposable income in each percentile.

Step 4: The welfare weights are interpolated using the function y^{ν}, where y is the average disposable income in each percentile, and ν is a parameter. Different values of ν lead to different welfare weights.

Step 5: We re-normalize the welfare weights so that they sum to 1.

Step 6: We plot the re-normalized welfare weights against the percentiles of the income distribution.

E.3 Construction of Likelihood of Registered Voters

We provide a detailed description of the methodology used to construct the likelihood of Social Architects being registered voters. The data on voting behavior and demographics are obtained from the Cooperative Election Study (CES) 2022 wave, which includes a nationally representative sample of 60,000 individuals.

Step 1: Using the CES data, we 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: We use the logit model 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 the Tax Formula

We present the details on the calibration of optimal labor income taxes. Data on the income distribution was obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). To simplify the analysis, we consider all individuals in the data as single filers, disregarding their tax filing status or whether they filed taxes at all. Our chosen measure of income is pre-tax labor income (*plinc*).

Step 1: We exclude individuals with negative labor incomes from our analysis.

Step 2: We discretize the data by computing the average labor income in various percentiles (1, . . ., 99, 99.9, 99.99).

Step 3: We convert this discrete income distribution into a smooth distribution in two steps. First, using the discretized income distribution, we estimate a smoothing splines regression, regressing the log of the average incomes in each percentile on the percentiles. The smoothing parameter is selected using cross-validation. Second, we use the predictions from the smoothing spline to create a smooth income distribution with 1000 equally spaced grid points (percentiles ranging from 0.1, 0.2, ... 99.9, 100).

F Proofs

F.1 Incentive Compatibility of the Staircase Procedure

In this section, we show the incentive compatibility of the staircase procedure. In each question that the Social Architects face, they have to choose whether to implement a Variable Reform (VR) or a Constant Reform (CR). The VR takes away \$t from the higher-income Recipient and gives \$pt to the lower-income Recipient. The CR takes away \$500 from the higher-income Recipient and gives \$500 to the lower-income Recipients. The fourth (last) question that Social Architects answer allow us to identify their welfare weights. If the Social Architect is randomly selected in the study, one of their decisions in one randomly selected question will be implemented.

Consider a simple version of the staircase procedure depicted in Figure A13. In the first question, a Social Architect decides between VR $(\frac{A+C}{2}, -\frac{B+D}{2})$ and CR $(500, -500)$. If the Architect chooses CR, then the following question asks them to choose between $VR(C, -D)$ and $CR(500, -500)$. If the Architect chooses VR, then the following question asks them to choose between $VR(A, -B)$ and $CR(500, -500)$. By construction, $\frac{A+C}{2} > 500$ and $\frac{B+D}{2} > 500$. Social Architects' decision in the second questions allows us to infer their welfare weights. The staircase is constructed in such a way that $C > A$ and $B > D$. This implies that for any arbitrary non-zero weights (g_1, g_2)

$$\begin{aligned} & g_1 C > g_1 A \quad \text{and} \quad g_2 B > g_2 D \\ \Rightarrow & g_1 C - g_1 A > g_2 D - g_2 B \end{aligned} \tag{A2}$$

We can define a *profile* of choices as the set of choices made across questions. The staircase procedure is incentive compatible if an individual with some welfare weights has no incentives to deviate from the profile that generates the exact same welfare weights. This can be reformulated as

Statement: There exists a unique profile for each set of weights (g_1, g_2)

We show this on a case-by-case basis. In each case, we consider two profiles of choices, and show that the two profiles cannot be rationalized by a given set of weights. This would imply that there exists a unique profile for each set of weights.

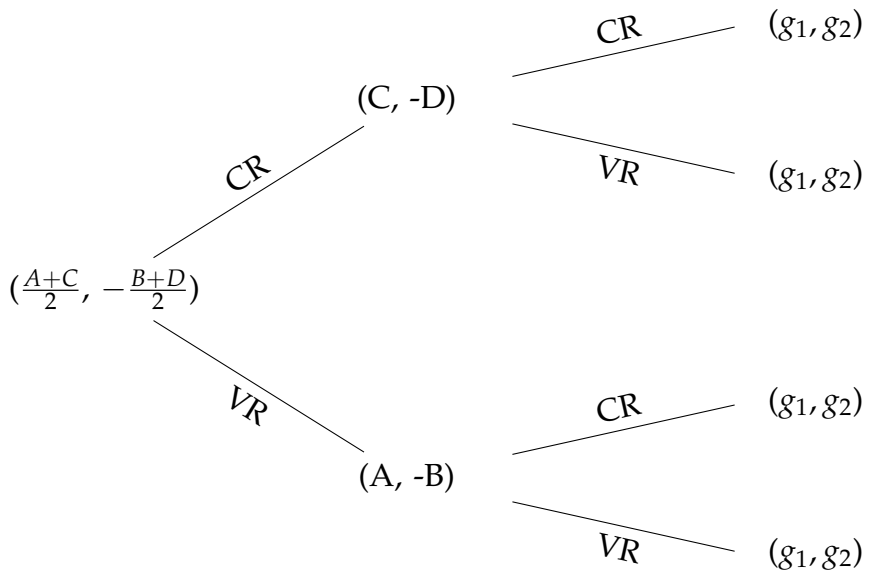


Figure A13: Simple Version of the Staircase Procedure.

Notes: VR and CR indicate that the Variable Reform and the Constant Reform was chosen in the previous question, respectively.

Case 1

Suppose a Social Architect chooses CR when deciding between CR and $(C, -D)$ and VR when deciding between the CR and $(A, -B)$. Choosing CR when deciding between the CR and $(C, -D)$ implies that $g_1 500 - g_2 500 > g_1 C - g_2 D$. Choosing VR when deciding between CR and $(A, -B)$ implies that $g_1 500 - g_2 500 < g_1 A - g_2 B$. Putting these two equations together, we have that $g_1 A - g_2 B > g_1 C - g_2 D$ or $g_1 C - g_1 A < g_2 D - g_2 B$, which contradicts Equation (A2). Thus, there does not exist a set of welfare weights (g_1, g_2) that can rationalize the choices in this case.

Case 2

Suppose a Social Architect chooses CR when deciding between CR and $(C, -D)$ and chooses CR when deciding between CR and $(A, -B)$. These two choices imply that $g_1 500 - g_2 500 > g_1 C - g_2 D$ and $g_1 500 - g_2 500 > g_1 A - g_2 B$, respectively. If we sum up the two inequalities, we obtain the following inequality

$$2g_1 500 - 2g_2 500 > g_1(A + C) - g_2(B + D) \tag{A3}$$

Now, assume that the Architect Chooses the VR in the first question when deciding between CR and $(\frac{A+C}{2}, -\frac{B+D}{2})$. This implies that $g_1 500 - g_2 500 < g_1 \frac{A+C}{2} - g_2 \frac{B+D}{2} \Rightarrow 2g_1 500 - 2g_2 500 < g_1(A + C) - g_2(B + D)$, which contradicts Equation A3. Thus, there

does not exist a set of weights (g_1, g_2), which can rationalize the choices in this case.

Case 3

Suppose a Social Architect chooses VR when deciding between CR and $(C, -D)$ and VR when deciding between (500, −500) and $(A, -B)$. These choices imply that $g_1 500 - g_2 500 < g_1 C - g_2 D$ and $g_1 500 - g_2 500 < g_1 A - g_2 B$, respectively. If we sum up both inequalities we obtain the following inequality

$$2g_1 500 - 2g_2 500 < g_1(A + C) - g_2(B + D) \tag{A4}$$

Now, assume that the Social Architect chooses the CR in the first question when deciding between (500, −500) and $\frac{A+C}{2}, -\frac{B+D}{2}$. This implies that $g_1 500 - g_2 500 > g_1 \frac{A+C}{2} - g_2 \frac{B+D}{2} \Rightarrow 2g_1 500 - 2g_2 500 > g_1(A + C) - g_2(B + D)$, which contradicts Equation A4. Thus, there does not exist a set of weights (g_1 and g_2), which can rationalize the choices in this case.

Case 4

Suppose a Social Architect chooses VR when deciding between CR and $(C, -D)$ and chooses CR when deciding between CR and $(A, -B)$. This implies that $g_1 C - g_2 D > g_1 500 - g_2 500$ and $g_1 500 - g_2 500 > g_1 A - g_2 B$, respectively. Putting these two inequalities together, we have

$$g_1 C - g_2 D > g_1 500 - g_2 500 > g_1 A - g_2 B \tag{A5}$$

Now, assume that the Social Architect chooses VR in the first question when deciding between CR and $(\frac{A+C}{2}, -\frac{B+D}{2})$. This implies that $g_1 500 - g_2 500 < g_1 \frac{A+C}{2} - g_2 \frac{B+D}{2}$. Then, $g_1 C - g_2 D > g_1 \frac{A+C}{2} - g_2 \frac{B+D}{2} > g_1 A - g_2 B$ iff $\frac{g_1}{g_2} > \frac{B+D-1000}{A+C-1000}$. Given that by construction $A + C > 1000$ and $B + D > 1000$, we ensure that the $\frac{g_1}{g_2}$ is non-negative. Similarly, when the Social Architect chooses CR in the first question when deciding between CR and $(\frac{A+C}{2}, -\frac{B+D}{2})$, then $g_1 \frac{A+C}{2} - g_2 \frac{B+D}{2} > g_1 C - g_2 D g_1 A - g_2 B$ iff $\frac{g_1}{g_2} < \frac{B+D-1000}{A+C-1000}$. The two inequalities would lead to mutually exclusive welfare weights, implying that there is a unique set of weights that rationalizes the decision to choose either CR or $(\frac{A+C}{2}, -\frac{B+D}{2})$.

G Instructions - Wave 1

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

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 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 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 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.

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 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]

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

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.

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]

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)]

page break

[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)]

page break

[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]

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.]

H Instructions - Wave 2

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

Treatment Base

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 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 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 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.

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 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 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]

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	**$8,000**	**$70,000**

Question 1/4: Please choose your preferred alternative:

Person A: +$1000 Person C: -$1000 ○	Person A: +$500 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 not deserve their current income and do not need their current income
- deserve their current income but do not need their current income
- do not 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 not deserve their current income and do not need their current income

- deserve their current income but do not need their current income
- do not 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.

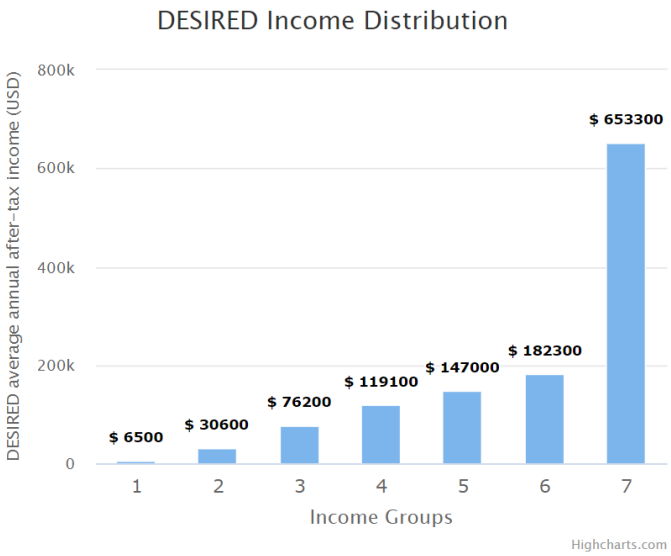
We would like you to indicate what you think the average tax rate for each tax group in the U.S. should be. This can be done as follows. **You can increase or decrease the average tax rates of the first six groups. The average tax rate of group 7 adjusts automatically so that all seven groups together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



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

page break

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.

Treatment Hypothetical

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

[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 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.

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.

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

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

Treatment Brackets

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

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

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**

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 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 Base. Decision Screens 1 to 6 are identical to the corresponding Decision Screens in Treatment Base, 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]

Treatment Self-Interest

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

[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 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 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 Brackets 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. 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* .

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.

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.