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

Francesco Capozza†
Krishna Srinivasan‡

April 27, 2023

Abstract

Implementing policy reforms involves weighing the gains for some individuals in society against the losses for others using welfare weights. We identify the welfare weights assigned by the general population of the U.S. to individuals in society 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 the social acceptability of reforms. We find that the welfare weights of the general population are slightly more progressive than the welfare weights implied by the existing policies and are much less progressive than the welfare weights commonly used in the optimal policy literature. We explore the implications for optimal labor income taxes.

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 settings such as income taxation, cash transfers, and in-kind transfers involve winners and losers. Evaluating the desirability of a reform requires assigning welfare weights to the winners and losers and subsequently weighing the gains of the winners against the losses of the losers. Welfare weights measure the value society places on providing an additional dollar of consumption to individuals. How should society weigh the gains of some individuals in society against the losses of others? The existing literature on optimal policies frequently assumes that the welfare weights assigned to individuals in society are inversely proportional to their incomes (e.g., Saez 2002). An alternative approach is to use the welfare weights implied by the existing tax schedule. Typically, these sets of welfare weights diverge from each other, complicating the selection of one over the other.[1] Furthermore, neither of these welfare weights may reflect societal preferences.

In this project, we elicit the welfare weights assigned by the general population of the U.S. using an experimental approach. 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 or the welfare weights implied by existing policies in the U.S. This is also a useful normative exercise since researchers and policymakers can use the elicited weights to evaluate reforms to ensure their social acceptability.

We conducted two real-stakes online experiments (N=4000) with a general population sample of 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, the Social Architects choose between various reforms. Social Architects’ decisions across these pairs reveal the welfare weights that they assign to the Recipients.

In our approach, the key information that the Social Architects have about the Recipients is their disposable incomes. We draw on the theoretical framework of Saez & Stantcheva (2016), who show that the welfare weights assigned based on the incomes of the Recipients are sufficient statistics for redistributive tastes. This means that the welfare weights are consistent with one or more underlying ideals that are part of a large set of

[1] There are also several limitations to using the welfare weights implied by the existing income tax schedule (see Stantcheva (2016) and Lockwood & 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 are sensitive to the assumptions about the elasticity of taxable income. Third, they can sometimes be negative, in which case, they cannot be used in policy formulas that only allow positive welfare weights.

ideals such as equality of opportunity, utilitarianism, distribution based on merit vs luck, or poverty alleviation. Two Social Architects with the same welfare weights may be guided by different ideals. However, their ideals, and consequently their welfare weights, have the same welfare and distributional implications. One key advantage of using the sufficient statistics approach is that the welfare weights assigned by the Social Architects can be directly plugged into optimal policy formulas.

We identify the estimates of ν such that the function y^{ν} provides a good approximation of the welfare weights assigned by the Social Architects to Recipients with various disposable incomes y, where ν is a parameter that governs the progressivity of the welfare weights. This function is commonly used in the literature (e.g., Saez 2002). Negative values of ν indicate progressive weights (weights decreasing with the incomes of the Recipients), while positive values of ν indicate regressive weights (weights increasing with the incomes of the Recipients). We find that the welfare weights assigned by the general population are characterized by $\nu \in (-0.3, -0.5)$. Since the elicited welfare weights are not entirely insensitive to the incomes of the Recipients ($\nu \approx 0$), the general population wants additional redistribution at the margin. The welfare weights of the general population can be directly used to evaluate reforms in which the welfare weights are conditioned on income levels. In other settings, our portable experimental method can be easily used to elicit the weights from a general population sample.

We compare the average welfare weights of the general population obtained from our experiment to the welfare weights implied by policies in the U.S. The weights implied by policies reflect politicians' aggregation of societal welfare weights and their political economy considerations. We find that the degree of redistribution implied by average welfare weights of the general population is about three to five times higher than the degree of redistribution implied by the income tax system in the U.S. ($\nu = -0.1$) computed by Hendren (2020), but similar to the degree of redistribution implied by a set of policies in the U.S. ($\nu = -0.26$) computed by Hendren & Sprung-Keyser (2020). Why explore several reasons why the general population weights are higher than the weights implied by the income tax system. We conjecture that the weights implied by the income tax system are lower because politicians aggregate the welfare weights of the general population by underweighting low-income individuals with progressive welfare weights and overweighting high-income individuals with less progressive welfare weights.

Next, we compare the average welfare weights of the general population to the welfare weights commonly assumed in the optimal policy literature ($\nu = -1$). Under the assumption of logarithm utilities, these weights correspond to utilitarian welfare weights. We find that the degree of redistribution implied by utilitarian welfare weights commonly

assumed in the optimal policy literature is about two to three times higher than the degree of redistribution obtained from the average welfare weights of the general population. Our results are consistent with the literature suggesting that the utilitarian ideal may not reflect societal preferences because it is far too progressive (Saez & Stantcheva 2016, Weinzierl 2014, 2017).

To understand the implications of the welfare weights, we calibrate the optimal labor income taxes in the U.S. that would result from different estimates of welfare weights. We use the tax formulas derived in Saez & Stantcheva (2016) and Saez (2001). We find that the average optimal marginal tax rates obtained by plugging in the upper bound of the general population weights is about 14 percentage points higher than the average optimal marginal tax rates obtained by plugging in utilitarian welfare weights typically used in the literature. The difference is 21 percentage points if we instead use the lower bound of the general population weights.

Can policymakers use the welfare weights of the general population to evaluate policies to ensure their social acceptability? To answer this question, we test if Social Architects' welfare weights predict their policy views. We find that Social Architects' assigned welfare weights predict their policy views on government redistribution and taxation of top-income earners. A benchmarking exercise reveals that the Social Architects' welfare weights predict their policy views as accurately as their stated political affiliation.

What does the experimental measure of welfare weights capture? Our experimental measure of welfare weights is designed to capture Social Architects' welfare preferences. In contrast, Social Architects' real-world welfare weights capture many factors that affect their preferences for redistribution at the margin. Some of these factors, which we measure in our study, include misperceptions about taxation and the income distribution and views about taxation and government. We elicit Social Architects' real-world welfare weights using a survey measure by asking them if, given the current incomes of individuals in society, incomes should be further redistributed. Using a covariate decomposition procedure proposed by Gelbach (2016), we find that only two factors affect Social Architects' experimental measure of welfare weights. These are Social Architects' beliefs about the externalities due to inequality (e.g., crime) and their beliefs about taxes on high-income individuals hurting the economy. On the other hand, we find that in addition to the two factors indicated above, several other factors, such as misperceptions about the level of taxes, misperceptions about the share of individuals earning less than $35,000, and beliefs about trickle-down economics affect Social Architects' real-world welfare weights. Our results suggest that using the experimental measure of welfare weights in policy applications may be desirable as it is less likely to capture factors orthogonal to welfare

preferences and may be less sensitive to misperceptions about relevant information.

We test if the variation in Social Architects' welfare weights is driven by welfarist motives or non-welfarist motives. This is an important exercise because the approach developed by Saez & Stantcheva (2016) allows for non-welfarist approaches in addition to the traditional welfarist approaches. In welfarist approaches, welfare weights depend on characteristics directly affecting Recipients' utility functions (e.g., Recipients' disability status). On the other hand, in non-welfarist approaches, welfare weights depend on characteristics that do not directly affect Recipients' utility functions (e.g., Recipients' parental income). Our results suggest that the Social Architects' welfare weights are guided both by welfarist and non-welfarist motives. Combining this finding with the finding that Social Architects' welfare weights exhibit considerable heterogeneity lends support to our general sufficient-statistics approach of allowing Social Architects to express various welfarist and non-welfarist ideals.

We find that Republican Social Architects have less progressive welfare weights relative to Democrats and independents, which is consistent with the results in the literature (e.g., Fisman et al. 2017, Singhal 2021). We show that 8% of the partisan gap in Social Architects' preferences for real-world redistribution at the margin is driven by their welfare weights. We find that 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). Our results imply that Social Architects' welfare preferences are guided, albeit to a small degree, by self-interest motives.

Our paper is related to four strands of literature. The first is the empirical literature that aims to understand the factors or ideals that influence people's preferences for redistribution (e.g., Cappelen et al. 2013, Drenik & Perez-Truglia 2018, Fong & Poutvaara 2019, Schokkaert & Devooght 2003, Weinzierl 2017). This literature typically uses treatments to identify if an ideal affects people's welfare preferences. For example, Drenik & Perez-Truglia (2018) shows that people's preferences for redistribution to low-income individuals depend on whether the low-income individuals are framed as being "lazy" or as "diligent." We take a different approach in our paper. We elicit people's welfare weights using a sufficient statistics approach that is general enough to capture various ideals but is unable to identify which ideals may be driving people's decisions. One advantage of our approach is that the elicited welfare weights can be directly used to evaluate policies.

The second is the literature that aims to elicit the welfare weights using surveys. Saez & Stantcheva (2016) elicit the welfare weights of a sample recruited from Amazon Mechanical Turk. The welfare weights elicited in their paper are only applicable to the linear income tax formula. In contrast, our experimental approach 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, explore whether people's welfare weights correlate with their policy preferences, and explore whether the general population weights are similar to the weights implied by the policies 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 allowing the utility functions of the Recipients to capture the Recipients' concerns (e.g., Fleurbaey & Maniquet 2006, Weinzierl 2014, 2018). For example, Weinzierl (2014) shows how the principle of equal sacrifice can be incorporated into the tax formulas. In contrast, our approach involves eliciting the welfare weights of the general population using a sufficient-statistics approach that can incorporate a large of ideals and using the aggregate welfare weights in optimal policy formulas. We find that Social Architects' welfare weights are heterogenous and are guided by welfarist motives and non-welfarist motives, which supports our general sufficient-statistics approach.

The fourth is the literature that identifies the social marginal welfare weights implied by the tax schedule of a country (e.g., Bourguignon & Spadaro 2012, Hendren 2020, Lockwood & Weinzierl 2016, Zoutman et al. 2013) or implied by a set of policies (Hendren & Sprung-Keyser 2020). In principle, these inverse-optimum weights can be used to evaluate other policies. However, there are several limitations with the inverse-optimum approach (see Stantcheva (2016) and Lockwood & Weinzierl (2016)). First, the inverse-optimum weights 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, the inverse-optimum weights are sensitive to the assumption about the elasticity of taxable income. Third, the inverse-optimum weights can sometimes be negative, in which case, they cannot be used in policy formulas that only allow positive welfare weights.

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 presents the results. Section 5 presents the discussion.

2 Theoretical Framework

This section presents a simple model adapted from Saez & Stantcheva (2016). The model helps understand the mapping from Social Architects' responses in the experiment to their welfare weights.

2.1 Setup

Consider a population (normalized to one) with a continuum of Recipients indexed

by j. The Recipients have the following money-metric (linear) utility function $u_j = z_j - T(z_j) - v(z_j; x_j^u, x_j^b)$, where $T(z_j)$ is the taxes paid as a function of income z_j and v is the dis-utility from work as a function of z_j, x_j^u, and x_j^b. The vectors x_j^u and x_j^b are characteristics that can influence Recipients' disutility from work, which we explain below. Taxes are set as a function of income only, such that Recipient j's consumption c_j is given by $c_j = z_j - T(z_j)$. There are no savings in this model.

A Recipient can decide to work more or less depending on the proposed taxes. The earnings choice of Recipient j, under tax system T, is given by $z_j(T)$. Consequently, the taxes paid by the Recipients and their disutility from work is now a function of $z_j(T)$. Recipient j's indirect utility function under tax system T is given by $U_j(T) = z_j(T) - T(z_j(T)) - v(z_j(T); x_j^u, x_j^b)$. With abuse of notation, T refers to the tax function as well as the tax system.

2.2 Generalized Social Marginal Welfare Weights

A Social Architect assigns generalized social marginal welfare weight to Recipient j under tax system T given by

$$g_j(T) = g(z_j(T) - T(z_j(T)), z_j(T); x_j^s, x_j^b). \tag{1}$$

Again, with abuse of notation, T refers to the tax function as well as the tax system. The welfare weight g_j measures the marginal value of consumption to Recipient j. The assigned welfare weights are relative and, therefore, defined only up to a multiplicative constant. The vector x_j^b includes a set of characteristics that affect Social Architects' welfare weights and Recipients' utilities (e.g., Recipients' disability status). The vector x_j^s contains a set of characteristics that only affect Social Architects' assigned welfare weights (e.g., Recipients' parental income). The vector x_j^u includes a set of characteristics that only affect the Recipients' utilities (e.g., Recipients' preferences for leisure). In the traditional "welfarist" approach, welfare weights depend only on x_j^b, while in the more recently explored "non-welfarist" approach, welfare weights depend only on x_j^s.

We assume that a Social Architect has preferences on the characteristics (x_j^s, x_j^b) that are relevant for redistribution, forms beliefs about the distribution of these characteristics across the income distribution, and assigns welfare weights based on these preferences and beliefs. For example, a Social Architect that only cares about equality of opportunity as an ideal would form beliefs about the share of individuals with low parental income across various incomes and would assign weights based on her beliefs.

In our setup, Social Architects can assign welfare weights based on characteristics that

are not observable to the policymaker or cannot be used to condition taxes on. The welfare weights are thus aggregated up to the level at which the policy is conditioned on (income z).[2] Recipient j earning z_j can be seen as an individual randomly selected from the group of individuals in society earning z_j. Let $h(z)$ be the earnings density. Then, the average welfare weight assigned to Recipient j is the average of the welfare weights assigned to all individuals earning z and is given by

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

The average welfare weights assigned to the Recipients are sufficient statistics for redistributive tastes. This means that the average weights can be consistent with one or more underlying ideals that are part of a large set of ideals, such as equality of opportunity, utilitarianism, distribution based on merit vs. luck, or poverty alleviation.[3] For example, a Social Architect who assigns welfare weights based on equality of opportunity and a Social Architect who assigns welfare weights based on the needs of individuals may have the same average welfare weights $\bar{g}_j(T)$. Even though these two Social Architects are guided by different ideals, their assigned welfare weights have the same welfare and distributional implications. A key advantage of using the sufficient statistics approach is that the estimated average welfare weights can be directly used to evaluate policies.

[2]There are two possibilities when policies are conditioned jointly on incomes and other observable characteristics. First, the welfare weights can be aggregated up to the level of the policy (jointly on incomes and other observable characteristics). Second, the welfare weights can be aggregated only up to the incomes of the Recipients. See Hendren & Sprung-Keyser (2020) for an implementation of the latter.

[3]See Saez & Stantcheva (2016) for an overview of ideals that can be incorporated by our approach.

2.3 Tax Reforms

Consider a simple setting where a Social Architect views two Recipients with incomes z_1 and z_2, such that $z_2 > z_1$. We consider a perturbed tax system $\tilde{T}$, such that $\tilde{T} = T + \epsilon R$ for a given ϵ. The "small" (marginal) reform R_j changes Recipients' tax liability, mechanically, by $\epsilon R = (\epsilon R_1, \epsilon R_2)$, but the reform does not affect their income because of the envelope condition. The reforms R involve taking money away from the higher-income Recipient and giving money to the lower-income Recipient. We can define the social welfare function of the Social Architects from tax system $\tilde{T}$, given the welfare weights evaluated at tax system T, as follows.

$$\begin{aligned} W(\tilde{T}|T) &= \bar{g}_1(T) \cdot U_1(\tilde{T}) + \bar{g}_2(T) \cdot U_2(\tilde{T}) \\ &= \bar{g}_1(T) \cdot (z_1(T) - T(z_1(T)) + \epsilon R_1 - v(z_1(T); x_1^u, x_1^b)) \\ &+ \bar{g}_2(T) \cdot (z_2(T) - T(z_2(T)) + \epsilon R_2 - v(z_2(T); x_2^u, x_2^b)). \end{aligned} \tag{3}$$

The welfare weights are evaluated at the initial tax system T and are thus fixed in Equation (3). Taking the first-order condition of Equation (3) with respect to ϵ, we get

$$\bar{g}_1(T) \cdot R_1 + \bar{g}_2(T) \cdot R_2 = 0. \tag{4}$$

The weighted sum of the reform, weighted by the welfare weights, has a first-order effect on the welfare of the Recipients. A reform R is desirable if and only if the weighted sum of the reform is greater than or equal to 0. Equation (4) provides the necessary conditions for a local optimum: if a tax system T is a local optimum, then the weighted sum of a reform is equal to 0.

2.4 Identifying Welfare Weights in the Experiment

In the experiment, Social Architects face pairs of Recipients, each of whom has an initial $C = \$1500$ bonus. We assume that the initial bonus C does not affect the Recipients' tax liability or behavioral responses. We also assume that the Social Architects' welfare weights are locally constant for the initial endowments we consider, i.e., $g(z_j(T) - T(z_j(T)), z_j(T); x_j^s, x_j^b) \approx g(z_j(T) - T(z_j(T)) + C, z_j(T) + C; x_j^s, x_j^b)$.

Social Architects are asked to Choose between a *Constant Reform*, which takes away \$500 from the higher-income Recipient and gives \$500 to the lower-income Recipient, and various *Variable Reforms*, which take away $\$t$ from the higher-income Recipient and give $\$pt$ to the lower-income Recipient. Our goal is to elicit the reform $(\$pt, -\$t)$ that

makes the Social Architect just indifferent between the reform $(\$pt, -\$t)$ and the reform $(\$500, -\$500)$.[4] If we equate the first-order condition in Equation (4) with $R = (\$pt, -\$t)$ to $R = (\$500, -\$500)$, we get

$$\tilde{g} = \frac{\bar{g}_2(T)}{\bar{g}_1(T)} = \frac{pt - 500}{t - 500}. \tag{5}$$

Equation (5) shows that the welfare weight assigned to Recipient 2 relative to Recipient 1 is an increasing function of the parameter p. If p is lower than 1, then the welfare weight attached to Recipient 2 is lower than the welfare weight attached to Recipient 1. Given that the income of Recipient 2 is higher than the income of Recipient 1, this implies progressive weights. If p is higher than 1, this implies regressive weights. If p equals 1, the weights attached to the two Recipients are the same. Since the welfare weights are only defined up to a multiplicative constant, Equation (5) is sufficient to obtain the (relative) welfare weights assigned to the two Recipients.

3 Experimental Design

Section 3.1 describes the key features of the design. Section 3.2 describes the treatments fielded in two waves of data collection. Section 3.3 describes the additional questions fielded in the two waves. Finally, Section 3.4 describes the implementation of the experiment. The full set of instructions for Wave 1 can be found in Appendix G and for Wave 2 in Appendix H.

3.1 Design Features

Participants in this study are either in the role of a "Social Architect" or "Recipient." The Recipients' only role is to receive money based on the Social Architects' decisions.

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 infer the Social Architects' welfare weights given the current tax and transfer system. In the experiment, we use the

[4] In principle, we could look for a reform $(\$pt, -\$t)$ such that the Social Architect is indifferent between the reform $(\$pt, -\$t)$ and no reform $(\$0, \$0)$. However, we wanted Social Architects to choose between different reform amounts so that they are less susceptible to status-quo bias.

word “after-tax income" to refer to the incomes of the Recipients.

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

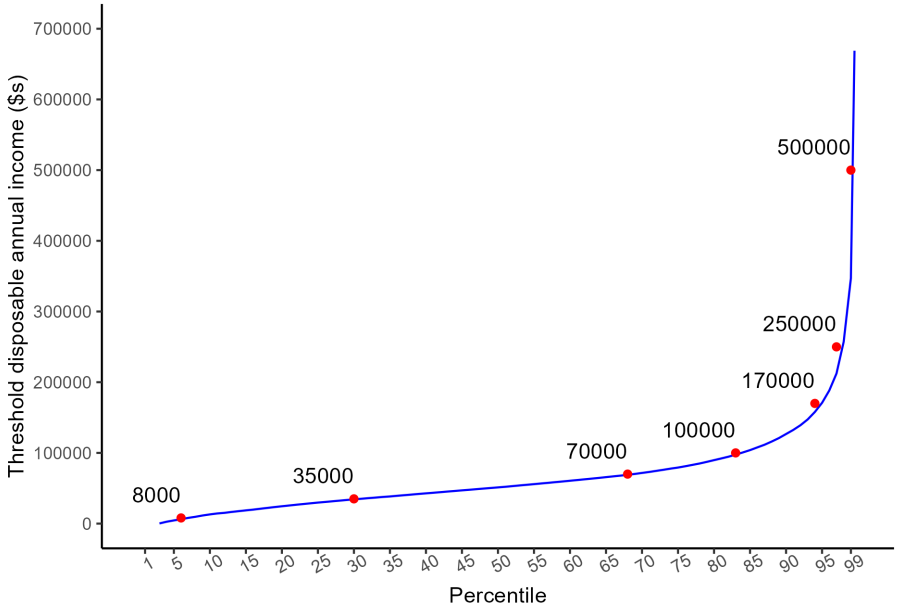


Figure 1: Disposable Incomes of the Seven Recipients

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 disposable annual income required for an individual to fall in the corresponding income percentile. Data on incomes 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 face six “decision screens." In each decision screen, they face a pair of Recipients. Table 1 lists the incomes of the Recipients in each decision screen. For a random half of the participants, the order of the decision screens is reversed.

Table 1: Pairs of Recipients in the Six Decision Screens

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

In each decision screen, Social Architects decide between a “Constant Reform" and various “Variable Reforms." A Constant Reform takes $500 from the higher-income Recipient

and gives \$500 to the lower-income Recipient. A Variable Reform reform takes \$t from the higher-income Recipient and gives \$pt to the lower-income Recipient.[5] Our goal is to elicit the value of p that makes the Social Architect just indifferent between the reform $(\$500, -\$500)$ and $(\$pt, -\$t)$. The smaller the value of the p, the less the Social Architect needs to give to the lower-income Recipient to be indifferent between $(\$500, -\$500)$ and $(\$pt, -\$t)$, implying that the Social Architect assigns a relatively higher weight on the lower-income Recipient.

Table 2 presents a set of fifteen questions with different Variable Reforms in each row. In each Variable Reform, the sum of the absolute value of the amounts is \$2000. Furthermore, higher row numbers in the table have Variable Reforms with higher values of p. Social Architects would typically choose a Constant Reform and then switch to a Variable Reform. Social Architects who switch to the Variable Reform later in the table (implying a higher p) have less progressive welfare weights. We identify the value of p by computing the mid-point of the Variable Reform in which Social Architects switch and the Variable Reform in the previous row. For Social Architects who switch in the first row (last row), the value of p is determined by computing the mid-point between the Variable Reform in the first row (last row) with the bound on the possible set of reforms.[6] We compute the relative welfare weights by plugging the reform amounts in Equation (5).[7]

Instead of presenting Social Architects with all the fifteen questions in each decision screen, we use the "staircase method" and present Social Architects with only four questions in each decision screen. All Social Architects start with the question 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

[5] In principle, we could look for a reform $(\$pt, -\$t)$ such that the Social Architect is indifferent between the reform $(\$pt, -\$t)$ and no reform $(\$0, \$0)$. However, we wanted Social Architects to choose between different reform amounts so that they are less susceptible to status-quo bias.

[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, which corresponds to their entire initial bonus. If a Social Architect always chooses the Constant Reform (never switch), 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, which corresponds to the amount in the Constant Reform.

[7] Suppose a Social Architect prefers a Constant reform in rows 1 through 6 and switches to the Variable Reform in row 7. This implies that the Social Architect is indifferent between (\$500, −\$500) and (\$912.5, −\$1087.5), where the latter was obtained by taking the mid-point of (\$875, −\$1125) and (\$950, −\$1050). Plugging these reforms into Equation (5), we get $\frac{\bar{g}_2}{\bar{g}_1} = \frac{pt-500}{t-500} = \frac{912.5-500}{1087.5-500} = 0.702$. This implies that the relative weight assigned to the higher-income Recipient is 0.702 times the weight assigned to the lower-income Recipient.

[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

Table 2: Set of Possible Questions in Each Decision Screen

Row	Constant Reform	Variable Reform
1	($500, −$500)	($550, −$1450)
2	($500, −$500)	($625, −$1375)
3	($500, −$500)	($700, −$1300)
4	($500, −$500)	($750, −$1250)
5	($500, −$500)	($800, −$1200)
6	($500, −$500)	($875, −$1125)
7	($500, −$500)	($950, −$1050)
8	($500, −$500)	($1000, −$1000)
9	($500, −$500)	($1050, −$950)
10	($500, −$500)	($1125, −$875)
11	($500, −$500)	($1200, −$800)
12	($500, −$500)	($1250, −$750)
13	($500, −$500)	($1300, −$700)
14	($500, −$500)	($1375, −$625)
15	($500, −$500)	($1450, −$550)

Notes: The table presents a set of fifteen questions from which four questions are selected in each decision screen. A 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.

Architects.

Overall, Social Architects face four questions in each of the six decision screens. Since there is a common Recipient (earning $70,000) across the six decision screens, Social Architects' decisions across the six decision screens allow us to recover the (relative) welfare weights assigned to each of the seven Recipients.

Figure 3 presents a screenshot of one of the questions used in the experiment. In the instructions, we do not use the words "Social Architect" or "Recipient." We refer to the Social Architects as "participants," and we refer to the Recipients as "Persons."

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. As seen in Figure 3, we encourage Social Architects to consider each question carefully since

representative samples across the world. 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 procedure.

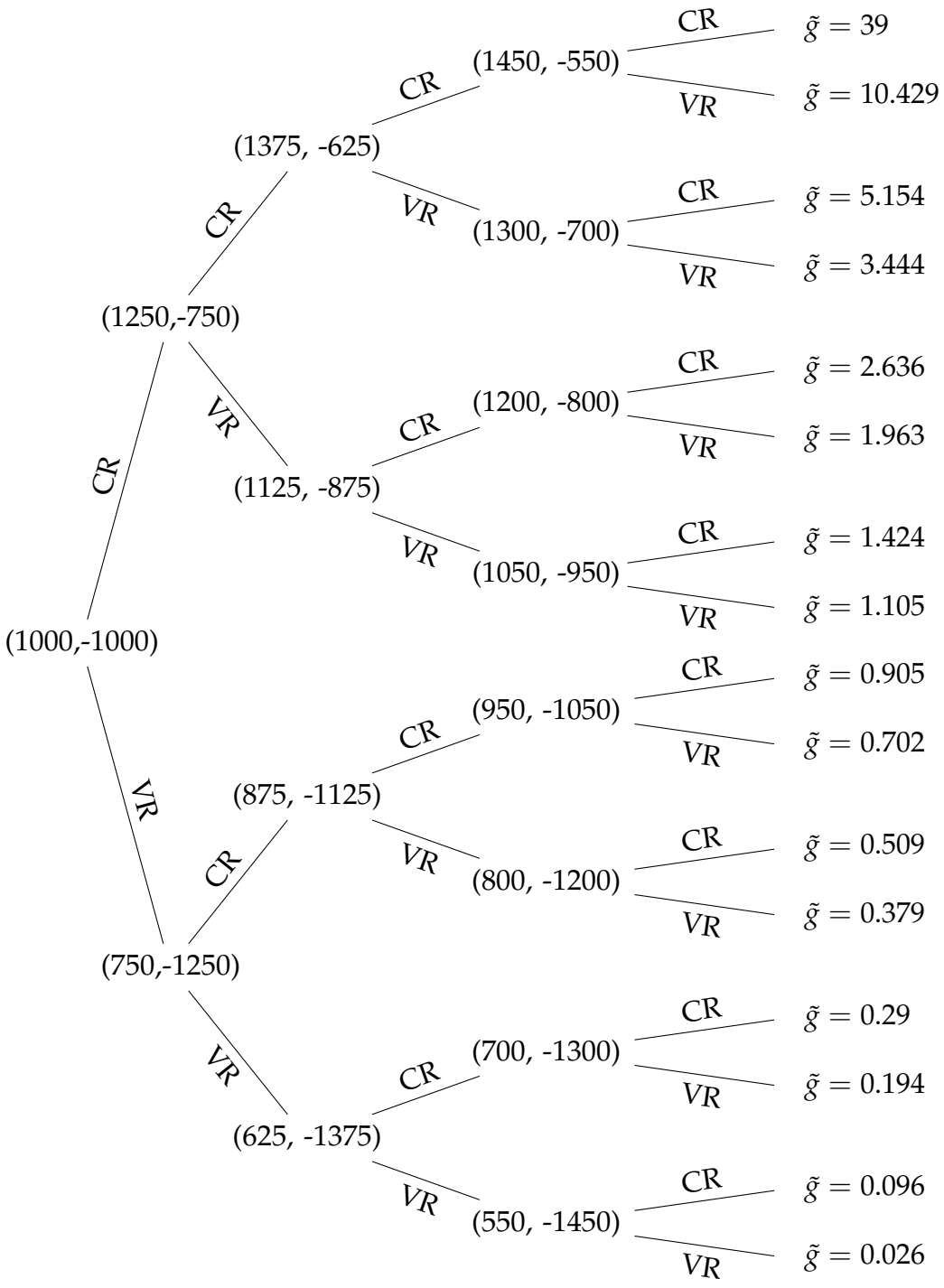


Figure 2: Variable Reforms Selected by the Staircase Method

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.

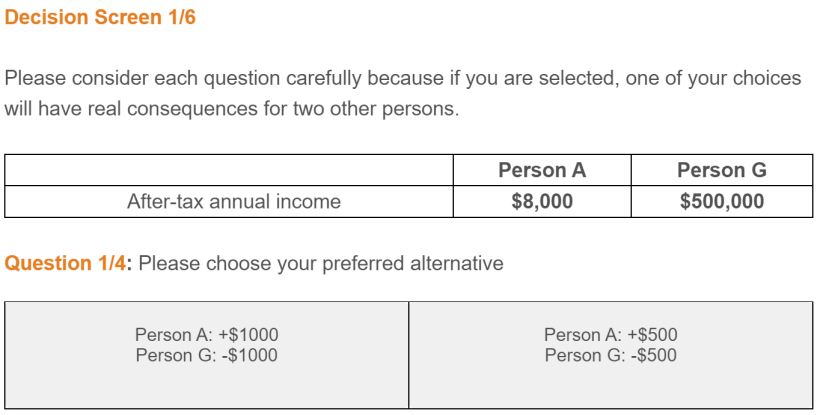


Figure 3: Screenshot of a Question Asked to Social Architects

one randomly selected question will have real consequences for two Recipients. Thus, 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.

3.2 Treatments and Waves

We implemented eight treatments across two waves of data collection to test the robustness of the welfare weights estimation with respect to changes in the experimental design. In Wave 1 of data collection, Social Architects are randomly assigned to one of four treatments in a 2×2 design. The design described above corresponds to Treatment Loss $\times$ Moderate. An overview of the treatments can be found in Table 3. The first dimension varies the framing of the reforms and helps test 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, Social Architects decide between $(\$500, -\$500)$ and $(\$pt, -\$t)$, and the weights are assigned to incomes that include an initial $1500 bonus. In Treatments Gain, Social Architects decide between $(\$2000, \$1000)$ and $(\$1500 + pt, \$1500 - t)$, and the weights are assigned to incomes that do not include an initial $1500 bonus. If we assume that the welfare weights are locally constant for the initial endowments we consider, i.e., $g(z_j(T) - T(z_j(T)), z_j(T); x_j^s, x_j^b) \approx g(z_j(T) - T(z_j(T)) + C, z_j(T) + C; x_j^s, x_j^b)$, the assigned weights should be the same in the two treatments.

The second dimension varies the Recipient that is common across the six decision screens, thereby affecting the income differences of the Recipients. This dimension allows us to test if Social Architects welfare weights are sensitive to the income differences of the Recipients. In Treatments Moderate (explained above), the Recipient common across

the decision screens has an income of $70,000. Since this Recipient is in the middle of the income distribution, the income differences between the Recipients is moderate. In contrast, in Treatments High, the Recipient common across the decisions screens has an income of $500,000. Since this Recipient is at the end of the income distribution, the income differences between the Recipients is High. Other than this, the two treatments are identical.

A treatment difference between Treatments Moderate and High can arise due to one of two reasons. First, a treatment difference may arise due to a behavioral effect. For example, when the Recipient common across decision screens has an income of $500,000, Social Architects may be insensitive to the differences in incomes of the other Recipients. Second, a treatment difference may arise due to a mechanical effect. In Treatments High, extreme choices (choosing the Variable Reform in every question or choosing the Constant Reform in every question) would result in the same welfare weights to Recipients one through six because we compare the first six Recipients to the seventh Recipient and because there is a bound on the welfare weights. Thus, there is a mechanical flattening out of the welfare weights. While there is also a mechanical flattening out of the weights in Treatments Moderate, the extent of the flattening out is lower since the Recipient common across the decision screens is in the middle of the income distribution, which allows the welfare weights to differ for the Recipients earnings below and above the Recipient common across the decision screens.

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 $\times$ Moderate in Wave 1. We include this treatment to test if there are differences across waves. One difference between the two treatments is that in Treatment Base, we stressed that the incomes of the Recipients were accrued after all taxes and transfers. Treatment Hypothetical is identical to Treatment Base except that the Social Architects make decisions regarding seven hypothetical Recipients and are informed that their decisions are purely hypothetical. This treatment allows us to test if Social Architects' welfare weights are sensitive to the existence of real stakes (albeit small stakes).

In the treatments mentioned above, Social Architects assume the role of impartial spectators, minimizing the influence of their self-interest motives. However, in reality, people are often affected by reforms themselves. To examine the role of self-interest motives, we include a treatment in which each Social Architect can potentially receive a bonus based on their decisions. Specifically, Social Architects view the income brackets of the seven recipients, who fall into seven income brackets spanning the disposable income distribution. Each Social Architect replaces the recipient whose income bracket matches

their own income.

In the above treatments, Social Architects assume the role of impartial spectators; their self-interest motives play a minimal role. However, in reality, people are often affected by reforms themselves. To examine the role of self-interest motives, we include a treatment in which each Social Architect can potentially receive a bonus based on their decisions. In particular, Social Architects view the income brackets of the seven Recipients, who fall into seven income brackets that span the disposable income distribution. Each Social Architect replaces the Recipient whose income bracket contains their own income. Additionally, We include Treatment Brackets, which is identical to Treatment Self-Interest, except that Social Architects act as impartial spectators and cannot receive a bonus based on their decisions. Comparing Social Architects in Treatment Self-Interest and Treatment Brackets helps explore whether Social Architects' self-interest motives affect their assigned welfare weights.

Table 3: Overview of Treatments

Wave 1		
Treatment	Framing of reforms	Income differences
Loss × Moderate	Loss	Moderate
Loss × High	Loss	High
Gain × Moderate	Gain	Moderate
Gain × High	Gain	High
Wave 2		
Treatment	Framing of reforms	Income differences
Base	Loss	Moderate
Hypothetical	Loss	Moderate
Brackets	Loss	Moderate
Self-interest	Loss	Moderate

3.3 Additional Questions

In this section, we present an overview of the additional questions faced by Social Architects. The questions can be found in Appendix Sections G and H. The definitions of the variables used in the analyses can be found in Appendix Section A.

Wave 1

We ask Social Architects two questions that elicit their views about 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 questions to test whether Social Architects' welfare weights predict their policy views.

Wave 2

We explore three sets of predictors of Social Architects' welfare weights. First, 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 explore whether Social Architects' welfare weights are influenced by welfarist motives (i.e., weights depend on the needs of the Recipients) or non-welfarist motives (i.e., weights depend on the characteristics of the Recipients that do not affect their needs). Second, 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 decisions 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.

We elicit Social Architects' real-world welfare weights using a survey measure designed to capture all the factors that influence their preferences for redistribution at the margin. We explore the factors that the experimental welfare weights and the real-world welfare weights capture. Some of these factors we measure in our study include misperceptions about taxation and the income distribution and views about taxation and government.

3.4 Implementation

For Wave 1, we recruited participants in the role of Social Architects from the data collection provider Lucid, which is commonly used in the literature (e.g., Haaland & Roth 2023, Haaland et al. 2021). The collected sample includes participants recruited based on quotas for sex, 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.

For Wave 2, we recruited participants in the role of Social Architects from the data collection provider Prolific, which is commonly used in the literature (e.g., Bursztyn et al. 2020, Enke et al. 2022). 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. For Wave 2, we used Prolfic instead of Lucid for two reasons. First, Prolific allows us to

pay participants directly, a feature required in Treatment Self-Interest and not available in Lucid. Second, using a different platform allows us to assess the robustness of the elicited welfare weights across platforms. We implemented the survey using the program oTree (Chen et al. 2016).

3.5 Pre-registration

We pre-registered the design as well as the analysis. We discuss the deviations from the pre-registration in Appendix Section B. There are no significant deviations in the design and implementation of the experiment. Many of the deviations in the analyses are minor. Some deviations from the pre-registered analyses involve the addition of new analyses. Other deviations involve the construction of variables.

4 Results

4.1 Summary Statistics

The data collection for Wave 1 began on 8 December 2021 and lasted approximately two weeks. We recruited a total of 6,735 participants. After dropping those with multiple responses, who do not consent to participate in the study, who do not pass the attention check, who do not pass the two comprehension checks, who do not reside in the U.S., and who fit into one of the demographic quotas that were full, we are left with 1,965 participants.[9]

The data collection for Wave 2 began on 14 December 2022 and lasted approximately six days. We recruited a total of 2,313 participants. After dropping those with multiple responses, who do not consent to participate in the study, who do not pass the attention check, and who do not pass the two comprehension checks, we are left with 1,992 participants.[10] All the sample restrictions were pre-registered, with the exception of dropping participants with duplicate responses.

[9] 0.09% of the participants have duplicate responses, 2.5% of the participants do not consent, 28.1% of the participants were dropped because the demographic quotas were full or because they do not reside in the U.S., 6.4% consent but drop out before the attention check, 28.2% fail the attention check, 16.5% pass the attention check but drop out before the comprehension checks, 21.3% fail the comprehension check questions, 5.8% drop out after passing the comprehension checks. 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 ($F = 1.763, p = 0.1522$).

[10] 0.3% of the participants have duplicate responses, 0.04% of the participants do not consent, 1.7% consent but drop out before the attention check, 2% fail the attention check, 3.1% pass the attention check but drop out before the comprehension checks, 5.9% fail the comprehension checks, 1.6% drop out after passing the comprehension checks. 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 presents the average characteristics of our sample relative to the characteristics of the population of the U.S. in 2019. The average population demographics are computed using the American Community Survey (ACS) 1-year estimates from 2019.[11] The population share of Republicans is taken from the Gallup poll.[12] Our sample characteristics from 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 reach our target sample size faster, our sample has a higher share of people with an education up to high school compared to the population. Our sample characteristics from Wave 2 also broadly match the population characteristics. However, there are several characteristics in which we do see large 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.

As pre-registered, we report analyses that are weighted using sampling weights (raking weights), which are constructed to ensure that the weighted averages of our sample characteristics match those of the population.[13] The average population characteristics are taken from Column (1) in Table 4.[14] When conducting analyses that only use the sample from Wave 2, the sampling weights are constructed such that the average sample characteristics of Wave 2 match the average population characteristics.

There are several differences between the two waves that are worth noting. First, the sample characteristics in Wave 1 more closely resemble those of the population because Wave 2 was conducted using Prolific, which has a limited set of quotas. Second, we find that Participants in Wave 2 were less likely to fail the attention check (28% fail in Wave 1 vs. 2% in Wave 2) or the comprehension checks (21% fail in Wave 1 vs. 6% in Wave 2), providing some suggestive evidence that the responses in Wave 2 may be more reliable. Our findings are consistent with the findings of Peer et al. (2022), who report a high data quality in Prolific. Finally, we coded Wave 1 using Qualtrics while we coded Wave 2 using oTree; participants in the two waves thus faced slightly different layouts during the survey.

[11] Note that the correct population averages listed in Table 4 are slightly different from those in the pre-registration document.

[12] To obtain the average population demographics, we restrict the ACS sample 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)

[13] The procedure creates raking weights based on a characteristic only if the sample 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.

[14] Note that the population characteristics indicated in Column (1) of Table 4 (the correct values) differ by a few percentage points from the estimates in the pre-registration document.

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 means are computed from the 2019 American Community Survey (ACS) 1-year estimates for the demographic variables. The Republican mean is obtained from Gallup Poll. 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 Lucid.

Table A1 in Appendix C.2 presents the average characteristics of our sample across the four treatments in Wave 1. We find that for several characteristics, the averages are statistically significantly different across treatments. However, with the exception of males and Republicans, the magnitude of the differences is small (less than five percentage points). Table A2 in Appendix C.2 presents the average characteristics of our sample across the four treatments in Wave 2. Again, we find that for several characteristics, the averages are statistically significantly different across treatments. With the exception of participants with a Masters's degree or above and participants residing in the western regions of the U.S., the magnitude of these differences is small. When exploring treatment effects, we

account for these imbalances by weighting each treatment to match the population.

4.2 Patterns in the Welfare Weights

In this section, we explore the patterns in the welfare weights assigned by the Social Architects to the Recipients. Figure 4 shows the average welfare weights (dots) pooled across all participants and the distribution of the weights (thickness of the shape around the dot). The average welfare weight assigned to the Recipient earning \$8000 is 0.22, which implies that Social Architects assign 22% of the total weight (100%) to the Recipient earning \$8000. The welfare weights are declining with the incomes of the Recipients, although not monotonically. The thickness of the figure around the dots indicates that there is considerable heterogeneity in the weights assigned to each of the seven Recipients.

We observe a relatively high weight assigned to the Recipient earning \$100,000 in treatments where the common Recipient across the decision screens has an income of \$70,000. This result is due to the fact that in the decision screen comparing the Recipient earning \$70,000 to the Recipient earning \$100,000, a share of libertarian participants who do not want to implement any of the Variable Reforms have regressive weights. These regressive weights pull up the average weight assigned to the Recipient earning \$100,000. We do not observe this pattern in the treatments where the common Recipient has an income of \$500,000. In the following sections, we explore the average weights of the general population by regressing the log of the welfare weights on the log of the incomes of the Recipients. This approach forces the fit of the average welfare weights to be strictly monotonic. Although this exercise ignores any non-monotonic patterns in the average welfare weights, it has the advantage of making the parameter estimates from our setting easily applicable to other settings.

We find that about 23% of the participants in our study have weakly monotonic welfare weights. Non-monotonic welfare 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 weights assigned to the poor, who they believe do not deserve additional money because they are considered "lazy" Drenik & Perez-Truglia (2018). To explore the progressivity of the Social Architects' welfare weights, we estimate the elasticity of Social Architects' welfare weights with respect to the income of the Recipients. In particular, we estimate the following regression for each Social Architect.

$$log(g(R_j)) = \beta_0 + \nu log(\textit{recipient income}_j) + \epsilon_j. \tag{6}$$

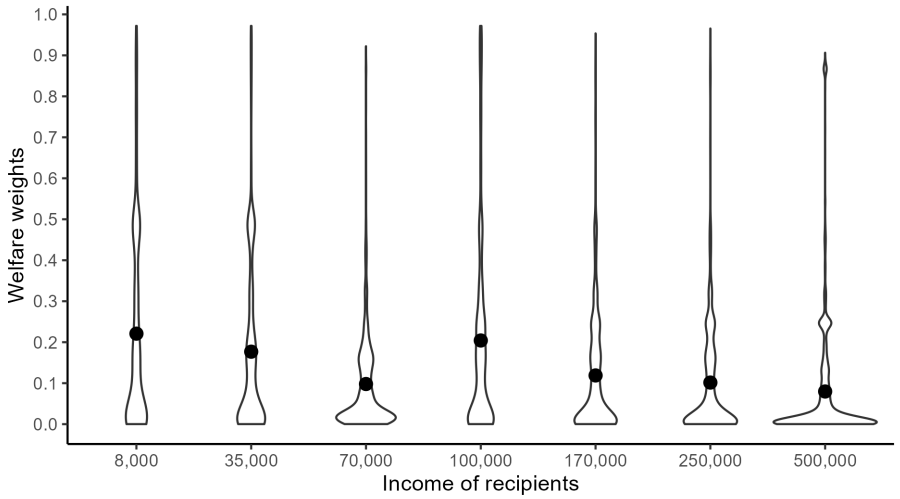


Figure 4: Welfare Weights Assigned by the General Population

Notes: The figure plots the average welfare weights (dots) assigned by the general population and the distribution of the weights (thickness of the shape around the dot) against the incomes of the seven Recipients.

Our findings suggest similar heterogeneity in the slope of the weights. Only about 15

The variable $g(R_j)$ is the welfare weight assigned by Social Architect i to Recipient j, *recipient income*$_j$ is the income of Recipient j, ν is the elasticity of a Social Architect's welfare weights with respect to the incomes of the Recipients. Negative values of ν indicate progressive weights (weights decreasing with the incomes of the Recipients), while positive values of ν indicate regressive weights (weights increasing with the incomes of the Recipients). Figure A1 in Appendix C.1 shows the cumulative distribution function of Social Architects' estimated elasticity parameters. Our findings suggest that around 66% of the Social Architects have progressive welfare weights, and 34% of the Social Architects have regressive welfare weights. 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 observe similar heterogeneity when we analyze the slope of the Social Architect's welfare weights with respect to the index of the Recipients instead (Figure A2 in Appendix C.1).[15] Table A3 in Appendix C shows the other patterns that we observe in the welfare weights. The table also shows the patterns separately by the two comparable treatments

[15]We present 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 range from \$8000 through \$500,000.

across waves: Treatment Loss × Moderate (Wave 1) and Treatment Base (Wave 2).

4.3 Treatment Effects

To explore the average welfare weights assigned by the Social Architects, we estimate the parameter ν that makes the function $(income)^{\nu}$ a best fit of the general population weights. Here, *income* refers to the incomes of the Recipients, and ν governs the progressivity of the welfare weights. This function is commonly used in the optimal policy literature (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). As seen in Table A3 in Appendix Section C, the range of ν is limited to $(-2.25, 2.25)$, which limits our ability to make uncover if people have extremely strong redistributive tastes. However, only 15% of the sample make choices that would result in welfare weights that are at the lower bound or upper bound, which implies that the limited range of ν would not severely affect the inferences we make. To compare the welfare weights of the general population to the function $(income)^{\nu}$, we estimate the following linear regression

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

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. As a notational shorthand, we define X as the vector of variables $x^1, ..., x^n$.

Figure 5 plots the coefficient estimates $\nu_0, .., \nu_n$ from the regression specified in Equation (7), where the vector X includes treatment dummies from Wave 1 and Wave 2. Since Social Architects' characteristics are not balanced across treatments, we present unweighted and weighted regressions. The weighted regressions are constructed by weighting each treatment based on the population averages. For the sake of brevity, we do not present the coefficient estimates $(\beta_1, ...\beta_n)$ of the treatment dummies. The coefficient estimate of $log(\textit{recipient income}_j)$ indicates the estimated elasticity of the weights with respect to the incomes of the Recipients in Treatment Loss × Moderate (the base category). The coefficient estimates $log(\textit{recipient income}) \times \textit{Treatment}$ can be interpreted as the change in the estimated elasticity in the given treatment relative to the elasticity in Treatment Loss × Moderate. Table A4 in Appendix Section C.2 presents the elasticity estimates separately by

treatments.

We find that the estimated elasticity of Social Architects' welfare weights in Treatment Loss $\times$ Moderate is -0.37. This result indicates that Social Architects in Treatment Loss $\times$ Moderate have "weak" redistributive tastes. The elasticity estimate implies that for every dollar the Social Architect gives to a Recipient, the Social Architect would give 0.63 cents to a Recipient earning twice as much.

Social Architects have more progressive welfare weights in Treatment Gain $\times$ Moderate (ν is 0.14 lower) relative to Treatment Loss $\times$ Moderate. In the treatments with a loss framing, Social Architects may be reluctant to take money away from the higher-income Recipient because they may believe that the Recipient may be worse off due to loss-aversion (Charité et al. 2015). Thus, the Social Architect may want to respect the reference point (status quo income) of the Recipient. In the treatments with a gain framing, loss-aversion is likely to play a smaller role since the reforms are framed as giving money to both the Recipient in the pair. While the estimated treatment effect is statistically significant, the effect size is modest, indicating that the welfare weights estimation is not very sensitive to whether we frame the reform as a gain or a loss.

Next, we compare Treatment Loss $\times$ Moderate to Treatment Loss $\times$ High. In the former, the common Recipient has an income of \$70,000, while in the latter, the common Recipient has an income of \$500,000. We find that Social Architects have less progressive welfare weights in Treatment Loss $\times$ High (ν is 0.27 higher). Part of this difference may reflect the fact that in Treatment Loss $\times$ High, Social Architects compare the Recipient earning \$500,000 to the other Recipients which may make them insensitive to the income differences between the other Recipients. This would lead to flatter weights in Treatment Loss $\times$ High. However, a part of this difference is a mechanical effect. In Treatment Loss $\times$ High, a Recipient who always chooses the Variable Reform in each decision screen would 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. Thus, there is a mechanical flattening out of the welfare weights in Treatment Loss $\times$ High. While there is also a mechanical flattening out of the weights in Treatment Loss $\times$ Moderate, the extent of the flattening out is lower. 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 $\times$ Moderate is -2.25, while it is -0.567 in Treatment Loss $\times$ High.

Treatment Loss $\times$ Moderate was implemented on Lucid while Treatment Base was implemented on Prolific. These two treatments are very similar otherwise.[16] We find that

[16]One other difference is that in Treatment Base, we emphasized that the incomes of the Recipients are

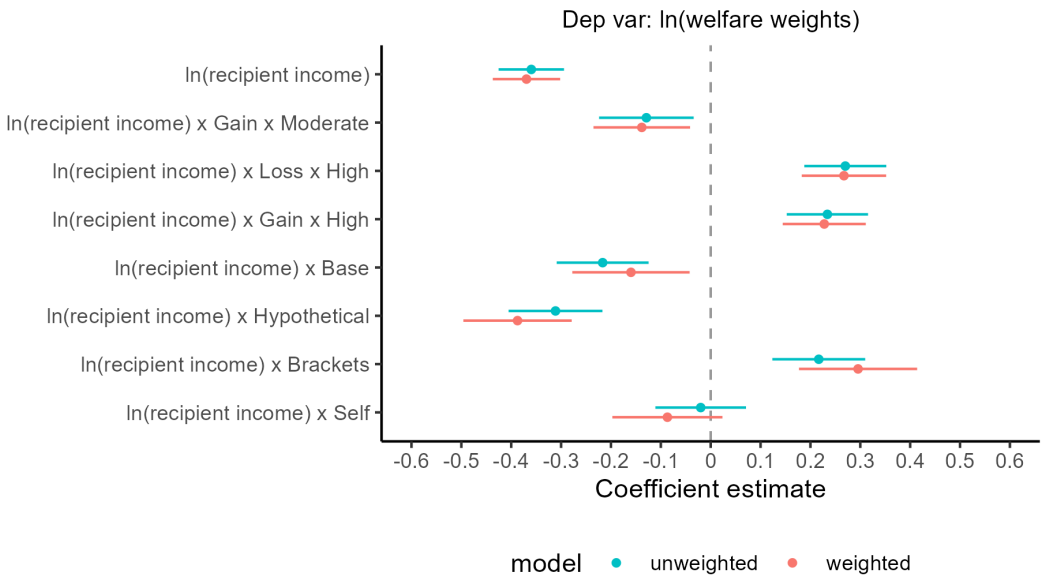


Figure 5: Welfare Weights and Treatments

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.

Social Architects are more progressive in Treatment Base in Wave 2 relative to treatment Loss $\times$ Moderate in Wave 2. However, the estimated effect is modest (-0.15). Table A3 in Appendix Section C.2 shows the patterns in the welfare weights across the two treatments. We find that Social Architects in Treatment Base are 2 percentage points ($p = 0.08$) less likely to choose the Constant Reform in every question and are 10 percentage points ($p < 0.01$) more likely to have weakly monotonically progressive welfare weights. These results may be driven by Social Architects being more attentive in Wave 2. However, we cannot rule out that Social Architects in Wave 2 have different preferences.

In Appendix Section , we explore other treatment differences. We explore the role of real-stakes by comparing Treatment Base to Treatment Hypothetical. We explore the role of framing the incomes of the Recipients by comparing Treatment Base to Treatment Brackets. Finally, we explore the role of self-interest by comparing Treatment Brackets to Treatment Self-Interest.

after all taxes and transfers. If this change has an effect, it is likely to make Social Architects in Treatment Base less progressive.

Discussion

Across the various treatments, we find that the value of ν is close to the value (-0.25) that the literature characterizes as "weak" redistributive preferences (Allcott et al. 2019, Saez 2002). This leads to the following result.

Result 1. *The average weights of the general population are characterized by "weak" redistributive tastes.*

Our result is also consistent with some studies in the literature that indicate that people have weak redistributive preferences in the U.S. (e.g., Weinzierl 2014, 2017). Since the estimated ν is not close to zero, people want some additional redistribution on the margin.

We identify the range of ν across treatments that can be used in the following sections. We exclude Treatment Loss $\times$ High and Gain $\times$ High because these treatments are less useful to identify Social Architects' weights since there is a mechanical flattening out of the weights in these treatments. We exclude Treatment Hypothetical because we conjecture that Social Architects in this treatment may not have considered the tradeoffs carefully. We exclude Treatment Brackets because in this treatment, we cannot cleanly control for Social Architects' beliefs about the incomes of the Recipients. Finally, we exclude Treatment Self-Interest since we are interested in Social Architects' welfare preferences unconfounded by self-interest motives. The elasticity parameters estimated using weighted regressions across our preferred treatments are $\nu \in (-0.37, -0.53)$. These estimates can be found in Table A4 in Appendix Section C.2.

4.4 Comparing to Inverse-Optimum Welfare Weights

In this section, we compare the average general population welfare weights to the weights implied by the income tax system in the U.S. and to the weights implied by a set of policies in the U.S. The welfare weights implied by the tax system and a set of policies likely reflect politicians' aggregation of societal welfare weights and their political economy considerations.

We obtain the welfare weights implied by the U.S. tax system from Hendren (2020). These taxes are computed using the universe of tax returns in 2012 and include ordinary income taxes, alternative minimum tax (AMT), earned income tax credits (EITC), state and local taxes, and Medicare. To obtain the "inverse-optimum" welfare weights, these taxes and other relevant parameters, such as the elasticity of taxable income, are plugged into the inverted optimal tax formula.

We obtain the welfare weights implied by a set of policies in the U.S. from Hendren & Sprung-Keyser (2020). Consider a policy that affects beneficiaries with incomes near

z^*. The Marginal Value of Public Funds (MVPF) of a policy is defined as the beneficiaries' 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 replicate 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 replicate the surplus accrued through the policy than through the tax schedule if and only if $s^* g(z^*) \geq c$. Re-writing this equation yields

$$MVPF = \frac{s^*}{c} \geq \frac{1}{g(z^*)}. \tag{8}$$

Thus, to obtain the inverse-optimum weights from a set of policies, we take the inverse of the MVPF of each policy. From the set of policies studied by Hendren & Sprung-Keyser (2020), we restrict the sample of policies to taxes, cash transfers, and in-kind transfers, which is the set of policies closest to the type of policies we study. Furthermore, we restrict the sample to policies with a positive MVPF since our theoretical framework and experimental design do not allow for a negative MVPF.

The welfare weights are sufficient statistics that can be considered "fixed." To compare different sets of welfare weights, we assume a parametric form for the welfare weights in which the welfare weights depend on the incomes of the Recipients. To compare the welfare weights, we compute the elasticity of the weights with respect to the incomes of the beneficiaries. We find that the elasticity of the inverse-optimum weights implied by the tax system is -0.10. The average weights of the general population (-0.37 to -0.53) are about 3.7 to 5.3 times more progressive than inverse-optimum weights. We find that the elasticity of the inverse-optimum weights implied by a set of policies is -0.27, which is much closer to estimates of the general population. This leads to the following result.

Result 2. *The general population welfare weights are similar to the weights implied by a set of policies in the U.S. but are more progressive than the weights implied by the U.S. income tax system.*

Figure 6 plots the average general population weights and the inverse-optimum weights against the 2019 disposable income distribution.[17] To compute the raw weights, we use the following function: (disposable income)$^{\nu}$, where ν represents the estimated elasticity discussed earlier. We then re-normalize the raw weights by dividing them by the

[17]We obtain the disposable income distribution data from Piketty et al. (2018), and we only include positive disposable incomes in our analysis.

sum of the raw weights so that they add up to 1.

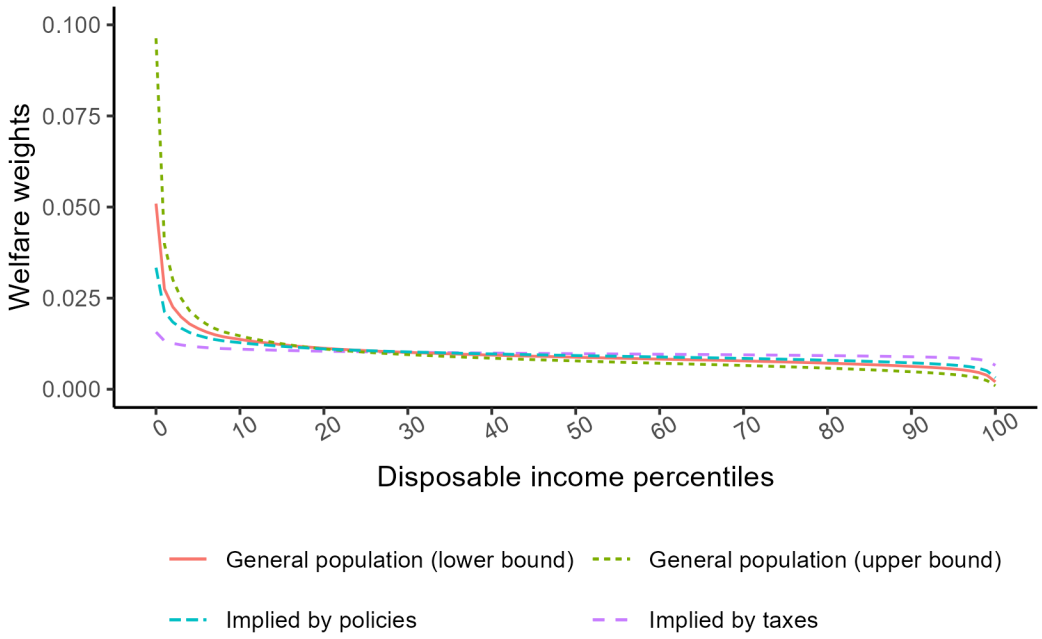


Figure 6: Average General Population Weights and Inverse-optimum Weights

Notes: The figure plots welfare weights against percentiles of the disposable income distribution. We use the function $(income)^{\nu}$ to interpolate the welfare weights for the income distribution and then re-normalize the welfare weights such that they sum to 1. The figure plots the re-normalized inverse-optimum weights implied by the tax system computed by Hendren (2020) ($\nu = -0.1$), inverse-optimum weights implied by a set of policies computed by Hendren & Sprung-Keyser (2020) ($\nu = -0.26$), lower-bound of the general population weights ($\nu = -0.37$), and upper-bound of the general population weights ($\nu = -0.53$). The disposable income distribution is obtained from Piketty et al. (2018) (*diinc*). We restrict the sample to positive disposable incomes.

What explains why the welfare weights of the general population are higher than the welfare weights implied by policies? First, we explore alternative ways of aggregating the welfare weights of the general population. Assuming that the Recipients' preferences are single-peaked, the politicians' aggregation of social welfare weights under the median voter theorem is equivalent to the median welfare weights. Thus, we compare the inverse-optimum weights to the median welfare weights of the general population. We compute the weighted median elasticity parameter of the Social Architects, weighting by the sampling weights. That is, we compute the median value of ν using Equation (6) pooling the data from all treatments. We find that the median value of ν is -0.33, which is still higher than the inverse-optimum weights.

Second, we explore alternative assumptions about the role of self-interest motives. Politicians may be aggregating the preferences of individuals who may be guided by self-

interest motives. In the treatments used above, self-interest motives play a minimal role by design. However, our study includes Treatment Self-Interest, which is better suited to elicit the welfare weights allowing for self-interest motives to play a role. In Section 4.3, we find that Social Architects' welfare weights are more progressive in Treatment Self-Interest relative to Treatment Brackets. This means that the general population welfare weights obtained from the other treatments would be more progressive if we allow for self-interest motives. Thus, allowing for self-interest motives does not make the inverse-optimum weights closer to the general population weights.

Third, we explore alternative assumptions about the median voter. There is some evidence that lower-income individuals in the U.S. are less likely to vote (Erikson 2015). In our experiment, we find that lower-income individuals have more progressive welfare weights. Thus, the aggregate welfare weights of the general population sample restricted to the sample that votes are likely to be closer to the inverse-optimum weights.

Fourth, we explore the hypothesis of "elite capture," namely that the politicians overweight the interests of the rich when aggregating societal welfare weights. In Appendix Section D.2, we find some evidence that Social Architects earning between \$210,000 to \$375000 assign a higher weights weight to Recipients with similar incomes as them, but limited evidence of self-interest motives playing a role for Social Architects earning above \$375,000. The aggregate general population weights obtained by overweighting the welfare weights of the higher-income individuals are likely to be closer to the inverse-optimum weights.

Overall, we find that the general population weights differ from the inverse-optimum weights, likely because the latter underweights the welfare weights of low-income individuals who do not vote and overweights the welfare weights of high-income individuals.

4.5 Comparing to Welfare Weights Used in the Literature

The optimal policy literature in public economics frequently assumes that the welfare weights are inversely proportional to the incomes of the Recipients. This corresponds to utilitarian welfare weights under the assumption of logarithmic utilities and can be represented by the function $(income)^{\nu}$, with $\nu = -1$.[18]

Figure A4 in Appendix Section C plots the general population welfare weights and the utilitarian weights. We find that utilitarian welfare weights assuming log utilities are 1.8 to 3.3 times more progressive than the welfare weights of the general population

[18]Suppose that in Equation 3, we assume that $U_j = log(z_j(T) - T(z_j(T)) - v(z_j; x_j^u; x_j^b)$ and that $\bar{g}_j = 1$, then in Equation (4), the welfare weights would be $1/(z_j(T) - T(z_j(T)) - v(z_j; x_j^u; x_j^b))$, which is the inverse of the disposable income.

($\nu \in (-0.37, -0.53)$). Our results are consistent with the literature suggesting that the utilitarian ideal may not reflect societal preferences because it is far too progressive (Saez & Stantcheva 2016, Weinzierl 2014, 2017).

Result 3. *Utilitarian welfare weights with logarithmic utilities commonly used in the literature are about 1.8 to 3.3 times more progressive than the average welfare weights of the general population.*

4.6 Calibrating Optimal Labor Income Taxes

We explore the optimal labor income taxes in the U.S. that result from different estimates of welfare weights. We use the tax formula derived in Saez & Stantcheva (2016) and Saez (2001), which is given as

$$T'(z) = \frac{1 - \bar{G}(z)}{1 - \bar{G}(z) + e(z) \cdot \alpha(z)} \tag{9}$$

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 driven by the substitution effect (people work less due to increased taxes). The formula assumes that there are 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 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 those earning above z. The optimal marginal tax rate is decreasing in $\bar{G}(z)$. To obtain $\bar{G}(z)$, we sum the welfare weights for all incomes above (including) z and divide the sum by $1 - H(z)$. The formula is given by

$$\bar{G}(z) = \frac{\int_{i:z_i \geq z} g_i \, d_i}{Prob(z_i \geq z) \cdot \int_i g_i \, di}. \tag{10}$$

We calibrate the sufficient statistics in the tax formula based on the literature and based on the welfare weights estimates provided in our paper. We set $\alpha(z) = 1.5$ (Piketty & Saez 2013) and set $e(z) = 0.25$ (Saez et al. 2012). Next, we calibrate $\bar{G}(z)$. We assign welfare weights using the function $g_i(z) = z^\nu$, where ν is a parameter that governs the progressivity of the weights. We re-normalize the weights such that they sum to 1 and

plug them in Equation (10). Different estimates of welfare weights can be obtained by plugging in different values for ν.

Figure 7 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.2.

We find that the slope of the optimal tax schedule obtained by plugging in the welfare weights implied by the tax schedule is very similar to the 2019 actual tax schedule. This helps us validate the exercise of obtaining optimal tax schedules. 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 14 percentage points higher than the average optimal marginal tax rates obtained by plugging in the weights typically used in the literature ($\nu = -1$). The difference is 21 percentage points if we instead use the lower bound of the general population weights ($\nu = -0.37$). Finally, we find that the optimal marginal tax rates corresponding to the general population weights are similar to those corresponding to the weights implied by a set of policies Hendren & Sprung-Keyser (2020) and slightly higher than those corresponding to the weights implied by the tax schedule Hendren (2020).

Result 4. *The optimal marginal tax rates corresponding to the welfare weights of the general population are lower than the optimal marginal tax rates corresponding to utilitarian weights but similar to the optimal marginal tax rates implied by a set of policies in the U.S.*

4.7 Do Welfare Weights Predict Policy Views?

We explore the empirical link between Social Architects' welfare weights and their policy views. This is an important exercise for two reasons. First, the existence of a link helps validate the weights as a measure of people's concerns for redistribution. Second, a strong link suggests that policymakers can use the welfare weights of the general population to evaluate policies to ensure their social acceptability.

In Wave 1 of data collection, we elicited Social Architects' policy views about increasing redistribution and the taxes on top-income earners. A modified version of the question on top-taxes was used by Cohn et al. (2019) and Kuziemko et al. (2015). The question on redistribution was used in the General Social Survey (GSS). The order of the questions is counterbalanced across participants.

In Appendix Section D.5, we explore the predictors of Social Architects' policy views.

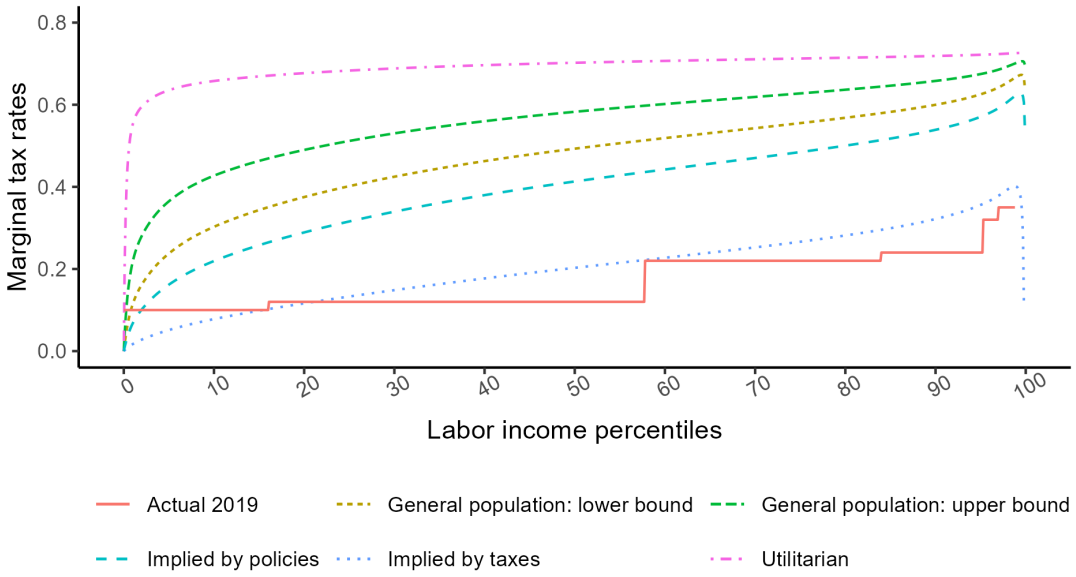


Figure 7: Marginal Tax Rates

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 tax system computed by Hendren (2020) ($\nu = -0.1$), inverse-optimum weights implied by a set of policies computed by Hendren & 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 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.

We find that Social Architects with more progressive welfare weights have more progressive views on top-taxes and redistribution. To understand how big a role welfare weights play in predicting Social Architects' policy views, we conduct a benchmarking exercise. In particular, we benchmark the predictive power of the elasticity of Social Architects' welfare weights against their stated political affiliation. We use a K-fold cross-validation procedure to assess the out-of-sample predictive power of the specifications. We find that the elasticity of Social Architects' welfare weights is just as good a predictor of their policy views as their stated political affiliation.

Result 5. *Social Architects' policy views can be predicted with similar accuracy using either their stated political affiliation or their assigned welfare weights.*

4.8 What Do the Welfare Weights Capture?

Our experimental measure of welfare weights is designed to elicit people's welfare preferences. However, people's real-world welfare weights may depend on several factors. In this section, we explore the factors that the experimental welfare weights and the

real-world welfare weights capture.

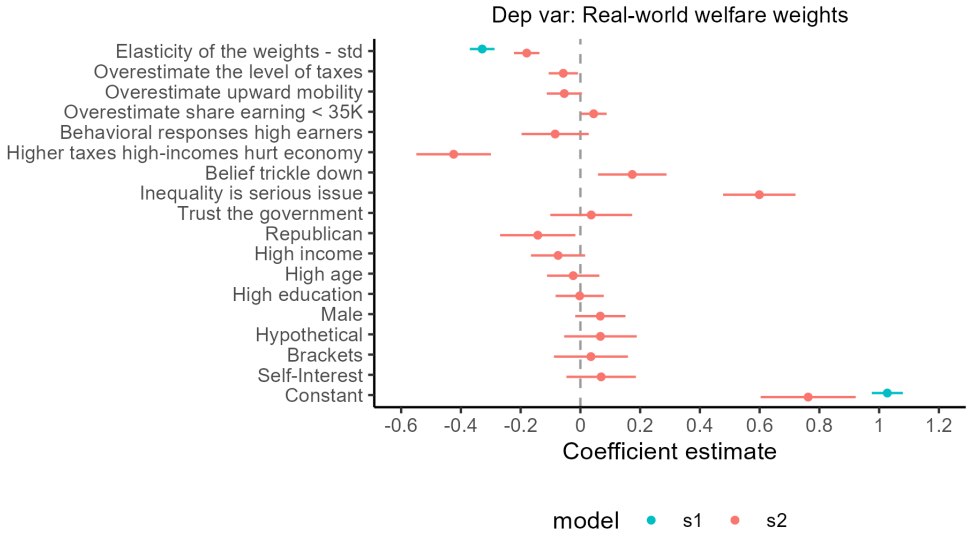
In Wave 2, we elicited Social Architects' real-world welfare weights using a survey measure. The question asks Social Architects to consider the current incomes of individuals in society after all taxes and transfers and asks them if they would like to redistribute incomes further. Their responses can 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.

We regress Social Architects' real-world welfare weights (elicited using the survey measure) on Social Architects' standardized elasticity of welfare weights (elicited using the experimental measure) and on various factors. These factors include a set of variables capturing Social Architects' misperceptions and views about taxation and government. The elasticity of Social Architects' weights is computed using Equation (6) and standardized to have a mean of 0 and a variance of 1. The explanation of the other variables can be found in Appendix Section A.1.

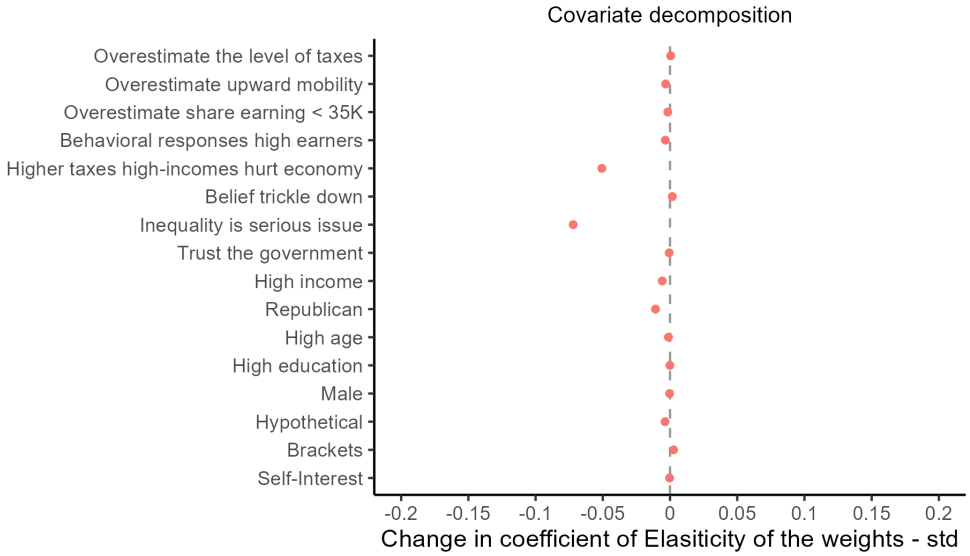
Figure 8a presents the coefficient estimates of the regression. In Model s1, we only include the elasticity of the weights as an explanatory variable, while in Model s2, we include all the variables indicated above as explanatory variables. In both models, we find that the elasticity of the weights is negatively correlated with the real-world welfare weights. An increase in the elasticity of the weights (less progressive weights) is associated with a reduction in the real-world weights (less progressive weights).

Figure 8a shows the direct effect of the variables of interest on Social Architects' real-world welfare weights. This is the effect not captured by our experimental measure of welfare weights. Social Architects who overestimate the level of taxes have less progressive real-world weights. Intuitively, Social Architects who think that society already bears a high tax burden would be 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 actually is, have less progressive real-world weights. 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 real-world weights. Social Architects' views about taxes and government also play a large role. Social Architects who think that higher taxes on high-income individuals hurt the economy have less progressive real-world weights. Social Architects who believe in trickle-down economics have more progressive real-world weights. However, this effect is counter-intuitive. Finally, Social Architects who believe that inequality is a serious issue because it can have externalities have more progressive real-world weights. Overall, we find that Social Architects' misperceptions and

views about taxation and government play a role in explaining their real-world welfare weights.



(a) Regression estimates



(b) Covariate Decomposition

Figure 8: Predictors of Real-World Weights and Experimental Weights

Notes: Panel (a) presents coefficient estimates. The dependent variable takes values from -2 to +2, with higher values indicating more progressive real-world welfare weights. 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 change in the coefficient estimate of *Elasticity of the weights - std* between Models s1 and s2 in Panel (a) that can be attributed to the other variables. The change is computed using the covariate decomposition procedure proposed by Gelbach (2016). Both figures use data from Wave 2.

We assume that the set of concerns captured by the real-world weights is a superset of the set of concerns captured by the experimental weights. Therefore, if a variable of interest indirectly affects Social Architects' real-world weights through the experimental weights, then this variable will change the coefficient estimate of their experimental welfare weights in Model s2. We interpret the change as the effect of the variable on the experimental weights. In Figure 8a, we observe that the coefficient estimate of the elasticity of the weights decreases by 50% in Model s2 when we include the variables of interest in the regression. Figure 8b presents the change in the coefficient estimate of the elasticity of weights (between Model s2 and s1) that can be attributed to each of the variables of interest. We calculate this change using the covariate decomposition procedure proposed by Gelbach (2016). 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. Interestingly, we find that many of the variables that directly affect Social Architects' real-world welfare weights (in Figure 8a) do not affect their experimental welfare weights.

Our results suggest that it may be desirable to use the experimental measure of welfare weights in policy applications insofar as they do not capture factors orthogonal to welfare preferences and are less sensitive to misperceptions.

4.9 Welfare Weights and Motives

We investigate whether Social Architects' welfare weights are driven by welfarist motives or non-welfarist motives. This is an important exercise because the approach developed by Saez & Stantcheva (2016) allows for non-welfarist approaches in addition to the traditional welfarist approach. In welfarist approaches, welfare weights depend on characteristics directly affecting Recipients' utility functions (e.g., Recipients' disability status). On the other hand, in non-welfarist approaches, welfare weights depend on characteristics that do not directly affect Recipients' utility functions (e.g., Recipients' parental income).

In Wave 2, Social Architects are asked about their beliefs regarding whether high-income individuals deserve and need their current income. Their answer options were as follows: (i) "do 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.3. 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 4.2) supports our general sufficient-statistics approach that can incorporate various welfarist and non-welfarist ideals.

Result 6. *Social Architects' welfare weights are guided by welfarist and non-welfarist motives.*

4.10 Individual Heterogeneity in Welfare Weights

To explore whether Social Architects' characteristics predict their assigned welfare weights, we estimate a version of Equation (7), where the vector X includes treatment dummies and Social Architects' characteristics. Model s1 in Figure 9 presents the results. In Models s2 to s6, the vector X in each model only includes one characteristic of the Social Architects. All models present estimates weighted using sampling weights.

We find that Republican Social Architects have less progressive weights (ν is 0.23-0.27 higher) relative to Democrats and Independents. Our results are consistent with findings in the literature indicating that Republicans have weaker redistributive tastes (Fisman et al. 2017, Singhal 2021). However, on average, Republicans also have progressive weights, implying that they want additional redistribution at the margin. Our results imply that a part of the partisan gap in preferences for redistribution is likely driven by the partisan gap in welfare preferences. We show in Appendix Section D.6 that Social Architects' welfare weights explain 8% of the partisan gap in preferences for redistribution at the margin. The partisan gap in preference is largely driven by concerns about externalities from inequality and concerns about the efficiency costs of taxation.

We find that Social Architects with higher incomes have less progressive weights (elasticity is 0.13-0.15 higher). This is consistent with some findings in the literature indicating that higher-income individuals have weaker redistributive preferences (Cohn et al. 2019, Singhal 2021). Appendix Section D.2 further explores how Social Architects' own income affects their welfare weights. We estimate a fixed-effects model to explore if Social Architects assign a higher weight to Recipients with similar incomes to their own relative to the other Recipients. We find some evidence that Social Architects assign a higher weight to Recipients with incomes similar to their own. However, we also find that this effect is larger for Social Architects with lower incomes.

Social Architects with a higher education level have more progressive weights (elasticity

is 0.08 lower), while Architects with a higher age have less progressive weights (elasticity is 0.08 higher). Social Architects' sex does not predict their welfare weights.

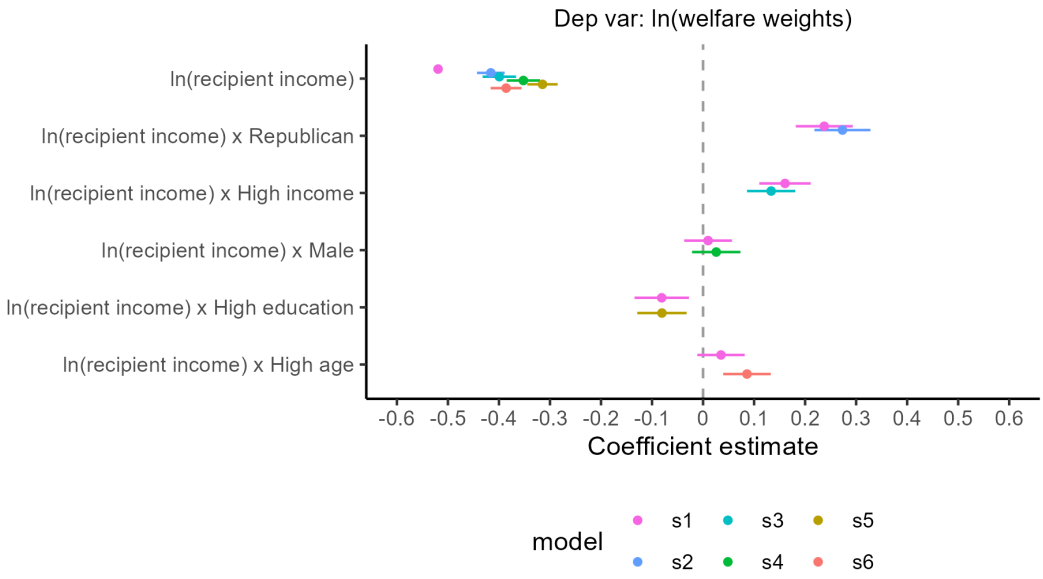


Figure 9: Social Architects' Welfare Weights and Characteristics

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. 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.

We conduct several other analyses, which we report in the Appendix. Table A5 in Appendix Section C presents several 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. We find similar results to the ones presented in Figure 9. We have the following result.

Result 7. *Social Architects with high incomes, low education, and a high age have less progressive welfare weights. Republican Social Architects have less progressive weights.*

5 Discussion

Various policy settings such as income taxation, cash transfers, and in-kind transfers require determining whether a reform is desirable. This involves weighing the gains to

some individuals in society against the losses to others using welfare weights. We elicit the welfare weights of the general population using a real-stakes online experiment in a general population sample of the U.S. population. In our experiment, participants in the role of "Social Architects" assign welfare weights to seven real "Recipients" with disposable incomes that span the income distribution of the U.S. The welfare weights assigned by a Social Architect to the seven Recipients are general enough to incorporate several different ideals that may guide Social Architects' decisions and can be directly plugged into optimal policy formulas.

Our results suggest that the average weights of the general population are more progressive than the welfare weights implied by policies in the U.S. but are less progressive than the welfare weights used in the optimal policy literature.

5.1 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 of the welfare weights. We also consider the median welfare weights. If preferences are single-peaked, the aggregate welfare weights according to the median voter theorem correspond to the median 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 & Bernheim (2021) in the context of ordinal preferences.

Second, it is unclear whether the aggregate welfare weights of the Social Architects should be used in policy formulas or whether the welfare weights of individual 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 to reduce concerns about measurement error and Social Architects' misperceptions.

5.2 Limitations of our Approach

Our sufficient-statistics approach to eliciting welfare weights, combined with the "small-reform" approach to taxation, has several limitations. First, the welfare weights assigned by the Social Architects may not be normatively appealing if they have inaccurate beliefs at various margins. However, there are two reasons why the mere presence of inaccurate beliefs may not be problematic. First, Social Architects may have misperceptions at various

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

Second, the welfare weights estimated in our paper cannot be used to evaluate non-marginal ('large") reforms.[19] 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.

Third, our approach is unsuitable in settings involving general equilibrium effects. However, this is a general limitation of the sufficient-statistics approach and extends to other objects in the tax formulas, such as the elasticity of taxable income.

5.3 Future Research

There are several avenues for future research. First, future work can test if the welfare weights estimated from our approach differ across time and countries. There is evidence in the literature suggesting that people's redistributive preferences may differ over time (Fisman et al. 2015) and across countries (e.g., Almås et al. 2020, Cappelen et al. 2013, Falk et al. 2018).

Second, future work can explore if people's welfare preferences depend on the policy domain. For example, people may want to assign different weights in the healthcare domain than in the education domain.

Finally, future research can estimate welfare weights for policies that are conditioned on Recipients' incomes as well as their characteristics ("tags"). The welfare weights estimated in our paper can only be applied to policies in which the welfare weights are conditioned on income levels. However, our general approach can be modified to allow welfare weights to be conditioned on Recipient's characteristics in addition to their income.

[19] 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. & Teso, E. (2018), 'Intergenerational mobility and preferences for redistribution', *American Economic Review* **108**(2), 521–554.

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

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

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

Bursztyn, L., Haaland, I., Rao, A. & Roth, C. (2020), Disguising prejudice: Popular rationales as excuses for intolerant expression, Technical Report 27288.

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

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

Chen, D. L., Schonger, M. & 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. & Smeets, P. (2019), 'Why do the rich oppose redistribution? An experiment with america's top 5%', *Working paper* .

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

Erikson, R. S. (2015), 'Income inequality and policy responsiveness', *Annual Review of Political Science* **18**, 11–29.

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

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

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

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

Fong, C. M. & 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.

Haaland, I. & 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. & 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. & 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. & Stantcheva, S. (2015), 'How elastic are preferences for redistribution? evidence from randomized survey experiments', *American Economic Review* **105**(4), 1478–1508.

Lobeck, M. & Støstad Nyborg, M. (2022), 'The consequences of inequality: Beliefs and redistributive preferences', *Working paper* .

Lockwood, B. B. & 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.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. (2022), 'Data quality of platforms and panels for online behavioral research', *Behavior Research Methods* **54**(4), 1643–1662.

Piketty, T. & Saez, E. (2013), Optimal labor income taxation, *in* 'Handbook of public economics', Vol. 5, Elsevier, pp. 391–474.

Piketty, T., Saez, E. & 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. & 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. & Stantcheva, S. (2016), 'Generalized social marginal welfare weights for optimal tax theory', *American Economic Review* **106**(1), 24–45.

Schokkaert, E. & 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. (2021*a*), Understanding economic policies: What do people know and how can they learn?, Technical report.

Stantcheva, S. (2021*b*), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 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. & 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 Design Details

A.1 Variable Definitions

A.1.1 Wave 1

Increase top-taxes: We ask the Social Architects whether the taxes levied on those in the top-income tax category should be increased, stay the same, or decreased. A similar question was used by Cohn et al. (2019) and Kuziemko et al. (2015). *Increase top-taxes* indicates Social Architects' responses (1 to 7) to the question, with higher values indicating more progressive views.

Increase redistribution: We ask the Social Architects whether the government should reduce income differences between the rich and the poor. This question is used in the General Social Survey. *Increase redistribution* indicates Social Architects' responses (1 to 7) to the question, with higher values indicating more progressive views.

A.1.2 Wave 2

Motives, Confidence, and Beliefs about the Source of Income

Motives: We ask the Social Architects 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 create three indicator variables from this question. The variable *High-income deserve but do not need* takes a value of 1 if a Social Architect indicates that high-income individuals deserve their current income but do not need their current income. The other two indicator variables are defined similarly. We also Social Architects a similar question about low-income individuals and define three indicator variables based on their responses. The base category corresponds to Social Architects who think that high-income individuals and low-income individuals do not deserve and do not need their current income.

Confidence: We ask Social Architects how confident they are that the choices they made in the welfare weight elicitation task reflect what they really think. *High confidence* is an indicator variable taking a value of 1 if a Social Architect's confidence about their choices reflecting what they really think is higher than the median and 0 otherwise.

Person rich due to effort: This is an indicator variable taking a value of 1 if a Social Architect indicates that a person is rich "Because she or he worked harder than others," and a value of 0 if a Social Architect indicates "Because she or he had more advantages than others".

Real-World Welfare Weights

Real-world welfare weights: We elicit Social Architects' real-world welfare weights. 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. *Real-world welfare weights* indicates Social Architects' responses ranging from -2 to $+2$, where positive values indicate that income should be further redistributed by taking from higher-income individuals and giving to lower/middle-income individuals, and negative values indicate that income should be further redistributed by taking from lower/middle-income individuals and giving to higher-income individuals. A value of 0 indicates that incomes should not be further redistributed.

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 explore perceptions along these four dimensions because they were the most predictive of people's policy views on redistribution in Stantcheva (2021*b*). For each of the four questions, Social Architects can select a number from 0 to 100 using a slider. 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 income tax = Beliefs about median income tax - 13

We take the true numbers from Stantcheva (2021*b*). We orient the gap in non-filers such that a lower gap in non-filers corresponds to an overestimation in the level of taxes. 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. They can select a number from 0 to 100 using a slider. We identify Architects' misperceptions by subtracting the truth (30) from their responses. We obtain the truth 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. They can select a number from 0 to 100 using a slider. We identify Architects' misperceptions by subtracting the truth (7.8) from their responses. We obtain the true 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 on 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 (2021*a*). The prompt for the question on inequality was taken from Lobeck & 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 ask 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: We ask Social Architects their beliefs about 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 very serious issue" or “A serious issue" and a value of 0 if a Social Architect indicates “Not an issue at all," “A small issue," or “An issue."

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 “Always" or “Most of the time" and a value of 0 for the responses “Only some times" or “Never."

B Pre-registration

The experimental design and the analyses were pre-registered.[20] We report three deviations from the pre-registration in the data collection and sample restriction. 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 (12 participants in Wave 1 and 14 participants in Wave 2). We did not pre-register this sample restriction as we did not expect participants to take the survey multiple times.

We report seven deviations from the pre-registration in the analyses. First, we construct the sampling weights based on the population estimates in Table 4 (the correct values). These estimates differ by a few percentage points from the estimates in the pre-registration document. Second, 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. Third, 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. Fourth, in the regressions that use a cross-section structure, we use the standardized elasticity of the welfare weights as our key variable. While we had pre-registered using the slope of the weights with respect to the Recipients' index as our key variable, we realize that this measure does not have an economic interpretation and is not easily transferable to other settings. Fifth, in the section (Appendix Section D.2) exploring the role of Social Architects' own income on their assigned weights, we present regressions in which the 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. Sixth, we explore the role of self-interest motives (Table A5 in Appendix Section C) by running a separate regression in each income bracket. We had

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

initially pre-registered an incorrect regression. Finally, the results in Table A6 were not pre-registered.

C Additional Tables and Figures

C.1 Additional Figures

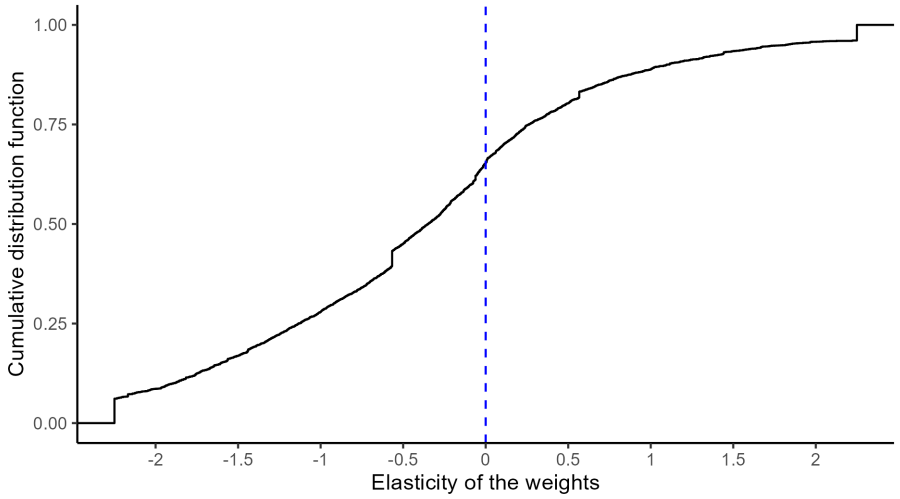


Figure A1: Distribution of the Elasticity of Social Architects' Welfare Weights

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.

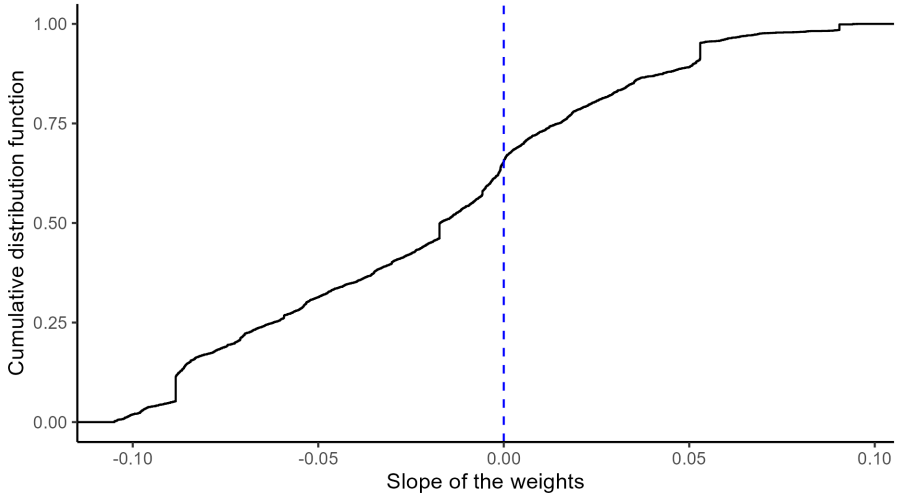


Figure A2: Distribution of the Slope of the Social Architects' Welfare Weights

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.

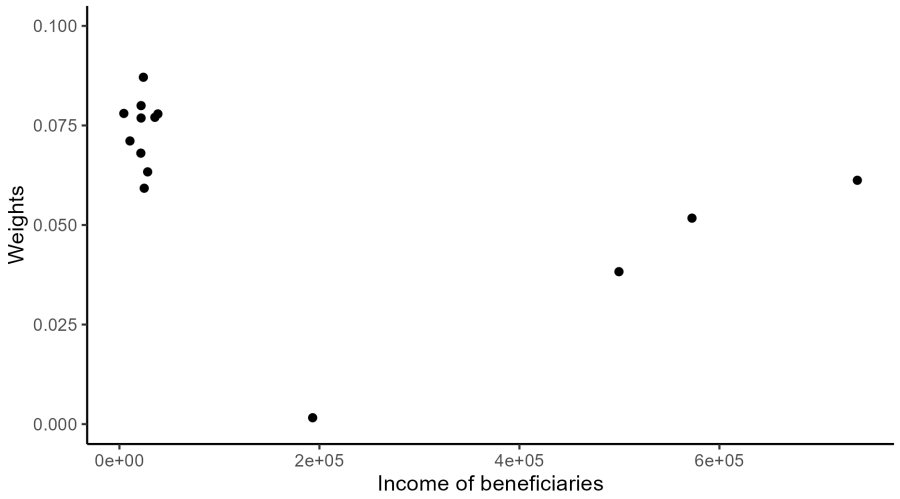


Figure A3: Inverse-optimum Weights and Incomes of Beneficiaries

Notes: The figure plots the welfare weights of various policies against the average incomes of the beneficiaries of the policies. The welfare weights are obtained from Hendren & Sprung-Keyser (2020) and are normalized to sum to 1.

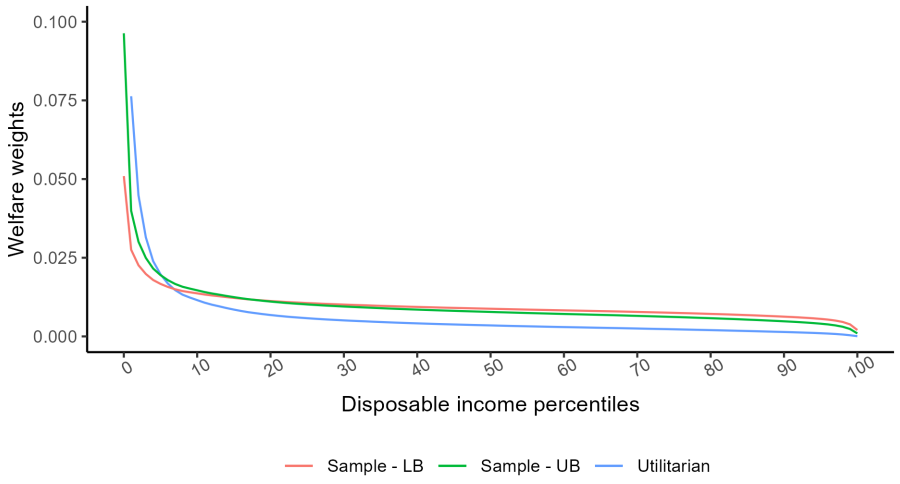


Figure A4: Average General Population Weights and Utilitarian Weights

Notes: The figure plots welfare weights against percentiles of the disposable income distribution. We use the function $(income)^{\nu}$ to interpolate the welfare weights for the income distribution and then re-normalize the welfare weights to sum to 1. The figure plots the re-normalized utilitarian weights ($\nu = -0.1$), lower-bound (LB) of the general population weights ($\nu = -0.37$), and upper-bound (UB) of the general population weights ($\nu = -0.53$). The disposable income distribution is obtained from Piketty et al. (2018) (*diinc*). We restrict the sample to positive disposable incomes.

C.2 Additional Tables

Table A1: Summary Statistics by Treatments in Wave 1

	Loss x High	Loss x Moderate	Gain x High	Gain x Moderate	p-value
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-value
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 (Base, Hypothetical, Brackets, Self-Interest) 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 Weights by Wave

Variable	Full Sample	Loss x Moderate	Base	p-value
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. 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 x Moderate and Treatment Base. *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 weights assigned to the Recipients are strictly decreasing (increasing) with the Recipients' income. Weakly progressive (regressive) weights imply that the weights assigned to the Recipients 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)
	mean	se	mean	se
loss_moderate	-0.36	0.03	-0.37	0.03
loss_high	-0.09	0.03	-0.1	0.03
gain_moderate	-0.49	0.03	-0.51	0.04
gain_high	-0.13	0.02	-0.14	0.02
T_base	-0.58	0.03	-0.53	0.05
T_hypo	-0.67	0.03	-0.76	0.04
T_brack	-0.14	0.03	-0.07	0.05
T_self	-0.38	0.03	-0.46	0.04
All	-0.35	0.01	-0.37	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 ν with respect to the incomes of the Recipients. Each row presents the estimates computed using the subset of Choice Architects indicated in the column "Case." Columns (1) and (2) report the estimates from an unweighted sample, while Columns (3) and (4) report the estimates from a weighted sample.

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 Moderate	-0.135*	-0.101	-0.125
	(0.079)	(0.086)	(0.079)
Loss x High	0.269***	0.303***	0.284***
	(0.061)	(0.069)	(0.062)
Gain x High	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. Columns (1) and (3) provide estimates after weighting the sample using population weights. The Standard errors are robust to heteroskedasticity (HC3).
*p<0.1, **p<0.05, ***p<0.01

D Additional Analysis

D.1 Treatment Effects

To explore the role of real-stakes, we compare Treatment Base to Treatment Hypothetical. In Treatment Hypothetical, Social Architects faced hypothetical Recipients. We find that participants' welfare weights are more progressive in Treatment Hypothetical (ν is 0.23 lower) relative to Treatment Base. Table A6 presents a series of regressions to explore the differences between the two treatments. We do not find any evidence that the share of participants always choosing the Constant Reform or always choosing the Variable Reform is different between the two treatments. While Social Architects spent 1.6 fewer minutes on the survey in Treatment Hypothetical, this difference is not statistically significant. Thus, the difference between Treatment Hypothetical and Treatment Base is not likely to have been driven by differential attention in the survey. One interpretation of the difference between Treatment Base and Hypothetical is that the existence of real stakes in Treatment Base makes participants think about the tradeoffs involved in real-world decisions more carefully.

Table A6: Comparing Treatments Base and Hypothetical

Dependent Variables:	Confidence	Time spent (mins)	Always CR	Always VR
Model:	(1)	(2)	(3)	(4)
Constant	4.366***	16.758***	0.039***	0.119***
	(0.058)	(1.542)	(0.011)	(0.026)
Hypothetical	-0.076	-1.679	0.018	-0.036
	(0.082)	(1.623)	(0.017)	(0.029)
Observations	997	997	997	997
Sampling Weights?	Yes	Yes	Yes	Yes

Notes: The table presents linear regressions. The sample is restricted to Social Architects in Treatments Hypothetical and Base. *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. *Confidence* is Social Architects' confidence that their decisions reflect what they really think. *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. In the regressions, each treatment is weights using sampling weights. The Standard errors are robust to heteroskedasticity (HC3).
*p<0.1, **p<0.05, ***p<0.01

To explore the role of framing of the incomes of the Recipients by comparing Treatment Brackets to Treatment Base. Treatments Brackets is similar to Treatments Base, with one

difference: we provide the Social Architects with income brackets of the recipients instead of their exact incomes. We assume that the Social Architects used the mean income of each bracket when assigning weights. These mean incomes are roughly similar to the incomes of the Recipients in Treatment Base, which we use to calculate the elasticities. Our findings show that Social Architects are less progressive in Treatment Brackets (ν is 0.45 higher) compared to Treatment Base.

To explore the role of self-interest motives, we compare the Social Architects in Treatments Brackets and Self-Interest. We find that participants in treatment Self-interest have more progressive weights than participants in Treatment Brackets. This indicates that self-interest motives do play a role. Figure A5 compares the assigned weights of participants in treatment Brackets and Self-Interest separately by income groups. We find that self-interest motives play a role for all income groups except the group earning above $135,000.

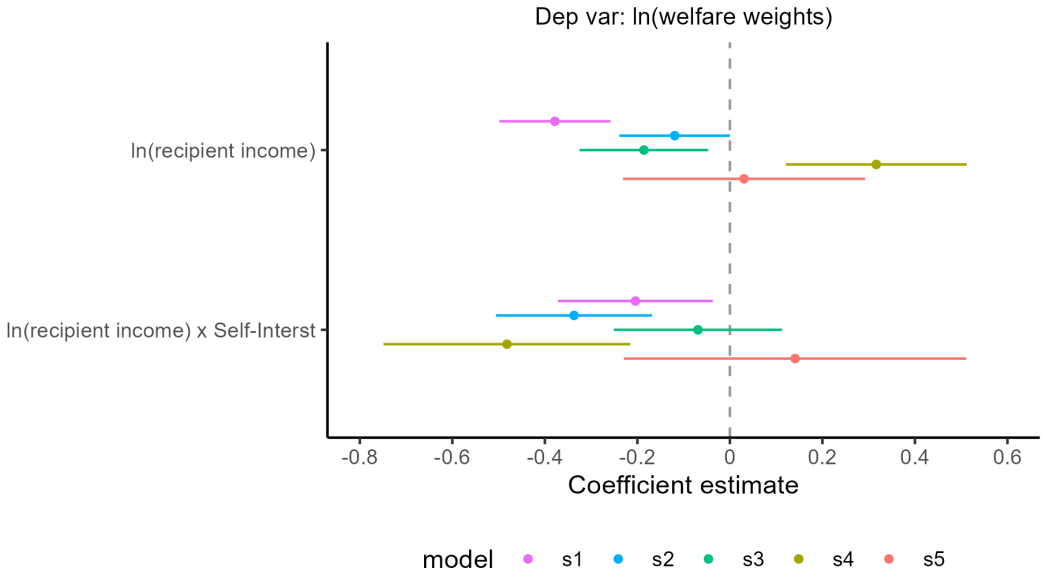


Figure A5: Social Architects' Weights and Self-Interest Motives

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 terms 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. In Model s1, the sample only includes Social Architects with incomes in the "$22,000 and below" bracket. In Model s2, the sample only includes Social Architects with incomes between $22,000 and $53,000. In Model s3, the sample only includes Social Architects with incomes between $53,000 and $85,000. In Model s4, the sample only includes Social Architects with incomes between $85,000 and $135,000. In Model s5, the sample only includes Social Architects with incomes in the following three brackets "$135,000 to $210,000," "$210,000 to $375,000," and "$375,000 and above." We do not present the main effects of *Self-Interest*. Error bars are computed using HC3 standard errors.

D.2 Role of Social Architect's Own Income

To explore the role that the Social Architects' own income plays in their assigned weights, we estimate the following regression

$$g(R_j)_i = \beta_1 \textit{Income near R1}_{ij} + ... + \beta_7 \textit{Income near R7}_{ij} + \gamma_i + \epsilon_{ij}. \quad (11)$$

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: (\$22,000, \$53000], (\$53,000, \$85000], (\$85,000, \$135000], (\$135,000, \$21000], (\$210,000, \$37500], (\$375,000, 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.

Figure A6 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. The figure presents error bars using HC1 standard errors since we could not compute HC3 standard errors in a weighted fixed-effects regression.

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 points 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.

Overall, given that only low-income and middle-income individuals assign a higher weight to Recipients with similar incomes, self-interest motives likely play a minimal role in our setting. In the regressions specified in Figure A6, we also include data from Treatment Self-Interest in which Social Architects' self-interest motives play a larger role. If we exclude Treatment Self-Interest in Figure A6, Social Architects' self-interest motives

would likely play a smaller role.

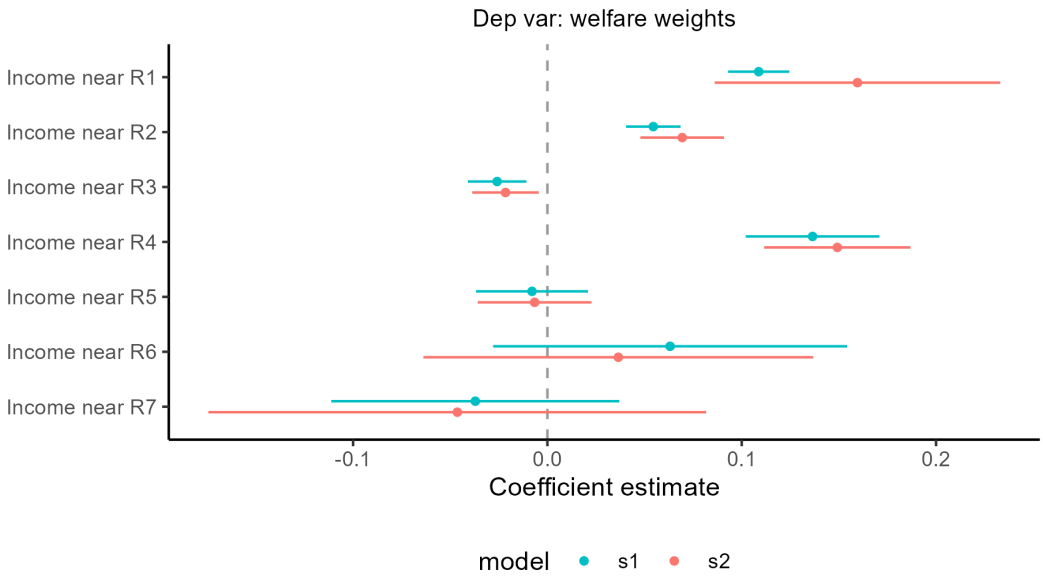


Figure A6: Weights Assigned to Recipients with Similar Incomes

Notes: The figure presents coefficient estimates. The dependent variable is the weight assigned by Social Architects. In Model s1, *Income near R1* takes a value of 1 if a Social Architect's income is less than or equal to $22,000. The other variables in Model s1 are similarly defined based on the following income brackets: ($22,000, $53000], ($53,000, $85000], ($85,000, $135,000], ($135,000, $210,000],($210,000, $37500], ($375,000, ∞). 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. Error bars are computed using HC1 standard errors.

D.3 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 A7 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 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.

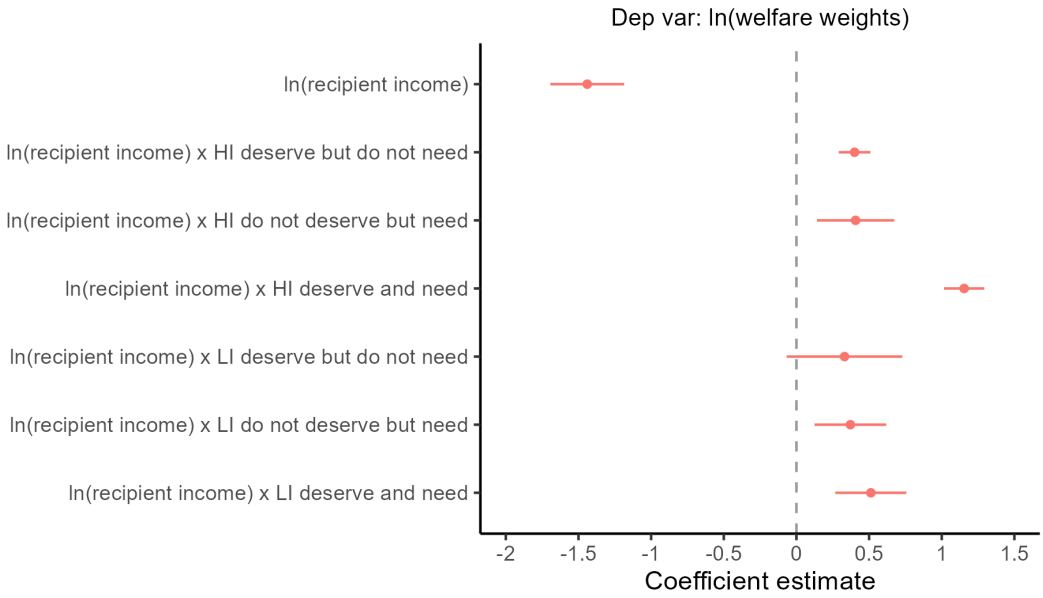


Figure A7: Social Architects' Weights and Motives

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. The other variables are defined similarly. 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 if the Social Architect indicates that 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.

D.4 Role of Beliefs About the Source of Income and Confidence

We explore whether Social Architects' beliefs about the source of income influence their welfare weights. 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 "Because she or he worked harder than others" and 0 if a Social Architect indicates "Because she or he had more advantages than others." Column (1) in Table A7 presents the results. 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.

At the end of six decision screens, we asked Social Architects how confident they were that the decision they made in the previous screens reflected what they really think. We explore whether Social Architects' confidence in their decisions influences 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. Column (2) in Table A7 presents the results. We find that Social Architects with high confidence have less progressive welfare weights relative to those with low confidence. However, the effect is not statistically significant.

Social Architects' responses to the question eliciting their confidence could range from "1: Not confident at all" to "5: Completely confident." We find that 83% of the Social Architects report a 4 or a 5, which indicates that most Social Architects' are confident that the decisions they made reflect what they really think.

D.5 Policy Views

We assess whether the Social Architects' welfare weights predict their policy views on top-taxes and redistribution. We elicited these policy views in Wave 1. The top-taxes question asks the Social Architects if they would like to increase the taxes on the top-income earners. The redistribution question asks the Social Architects if they want the government to do something to reduce income differences in society. For both questions, higher values indicate stronger redistributive tastes.

To understand the effect of the Social Architects' welfare weights and their characteristics on their policy views, we present two linear regressions. The dependent vari-

Table A7: Social Architects' Welfare Weights, Beliefs About the Source of Income, and Confidence

Dependent Variable:	ln(welfare weights)	
Model:	(1)	(2)
Constant	3.825***	1.856***
	(0.292)	(0.343)
ln(recipient income)	-0.668***	-0.478***
	(0.025)	(0.030)
Person rich due to effort	-9.731***	
	(0.662)	
ln(recipient income) × Person rich due to effort	0.872***	
	(0.057)	
High confidence		-1.132**
		(0.543)
ln(recipient income) × High confidence		0.075
		(0.047)
Observations	13,944	13,944
Sampling Weights?	Yes	Yes

Notes: The table presents coefficient estimates. The dependent variable is the log of the welfare weights assigned by Social Architects. In Column (1), the explanatory variables include 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 believes that high-income individuals are rich due to effort and 0 otherwise. In Column (2), the explanatory variables include the log of the incomes of the Recipients, *High confidence*, and the interaction terms 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 regressions are weighted using sampling weights. The regressions use data from Wave 2. The Standard errors are robust to heteroskedasticity (HC3).
*p<0.1, **p<0.05, ***p<0.01

ables in these regressions are *Increase top-taxes*, a variable taking values from 1 to 7, with higher values indicating a greater desire for the government to increase top-taxes, and *Increase redistribution*, a variable taking values from 1 to 7, with higher values indicating a greater desire for the government to do something to reduce inequality, respectively. The key explanatory variable is the standardized elasticity of Social Architects' weights.[21]. In particular, we estimate the following regression for each Social Architect

[21]Note that in the regressions in this section, we had initially pre-registered using the slope of the weights as the key explanatory variable.

$log(g(R_j)) = \beta_0 + vlog(recipient\ income_j) + \epsilon_j$, where $g(R_j)$ is the weight assigned by Social Architect i to Recipient j and $recipient\ income_j$ is the disposable income of Recipient j. We standardize the coefficient estimate $\hat{v}$ by subtracting the mean and dividing by the standard deviation. In these regressions, the other explanatory variables include a set of treatment dummies, Social Architects' characteristics, an indicator variable indicating the order of the decision screens, and an indicator variable indicating the order of two policy views questions.

Figure A8 presents the results. We find that the elasticity of Social Architects' weights are negatively correlated with their preferences to increase top-taxes and increase redistribution. This means that a Social Architect with more progressive welfare weights has a stronger preference to increase top-taxes and redistribution. Republicans and higher-income individuals have a weaker preference for increasing top-taxes and redistribution.

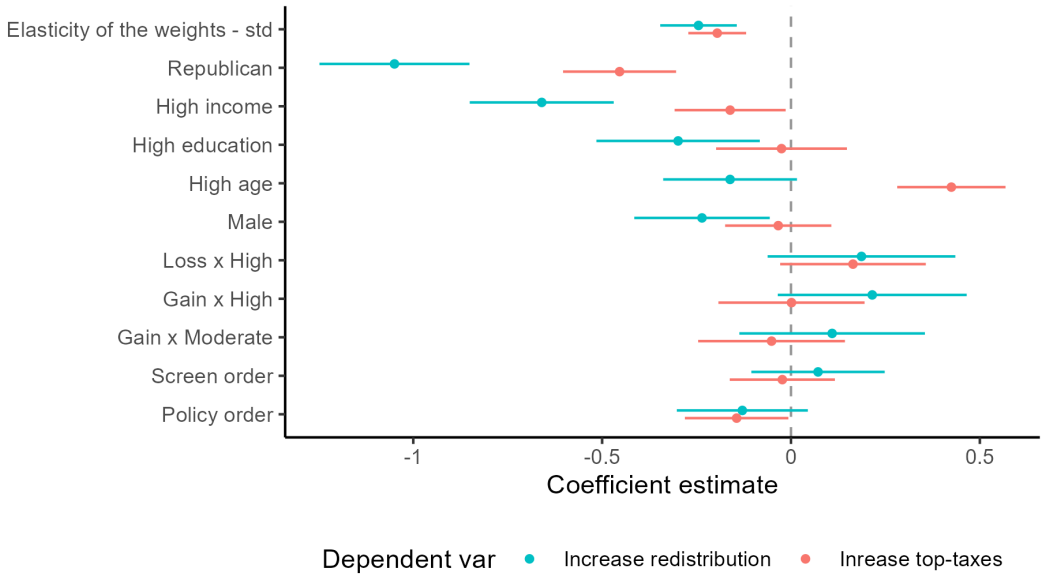


Figure A8: Predictors of Policy Views

Notes: The figure presents coefficient estimates. *Increase redistribution* 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 - std* is the standardized elasticity of Social Architects' weights with respect to the incomes of the Recipients. The other variables in the regression are *Republican* (=1 if Republican), *High Income* (= 1 if above median income), *Male* (=1 if male), *High Education* (=1 if above median education), *High Age* (=1 if above median age), *Screen Order* (dummy indicating the order of the decisions screens), and *Policy Order* (dummy indicating the order of the policy views questions). The regressions uses data from Wave 1.

To understand how big a role welfare weights play in predicting Social Architects' policy views, we conduct a benchmarking exercise. In particular, we estimate a series

of linear regressions in which the dependent variable is either Social Architects' policy views on top-taxes or their views on redistribution. The explanatory variable of interest is either the Social Architects' political affiliation or the Social Architects' elasticity of the welfare weights. To assess the predictive power of a specification, we compute the root mean squared error (RMSE) using a K-fold cross-validation procedure. [22] The K-fold cross-validation is a common method to assess the out-of-sample prediction of a specification. The lower the RMSE of a specification, the higher the predictive power of that specification.

Table A8: Predictors of Policy Views - 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	Increase redistribution	Republican	No	2.00
6	Increase redistribution	Elasticity of the weights	No	2.06
7	Increase redistribution	Republican	Yes	1.96
8	Increase redistribution	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 $k = 4$. *Increase redistribution* 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* takes a value of 1 for Republicans and a value of 0 for Democrats or Independents. The controls in the regression include a set of are *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), *Screen Order* (dummy indicating the order of the decisions screens), *Policy Order* (dummy indicating the order of the policy views questions). The regression uses data from Wave 1. Error bars are computed using HC3 standard errors.

Table A8 presents the RMSE from various specifications. Across the specifications, we find that Social Architects' political affiliation is just as good a predictor of Social Architects' policy views as their welfare weights. Using people's welfare weights to predict their policy views instead of using their political affiliation may be desirable in situations where eliciting people's political affiliation may affect their responses on other questions or in situations where there is social desirability bias.

[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 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 take the square root of the average of the squared errors across all four sub-samples.

Next, we explore how Social Architects' welfare weights assigned to the Recipients predict their views on top-taxes. We regress Social Architects' views on top-taxes on their welfare weights assigned to Recipients 1 through 7, excluding Recipient 4; We exclude the welfare weights assigned to Recipient 4 to prevent the problem of multicollinearity. In the regression, we also control for a set of treatment dummies, a dummy variable indicating the order of the decision screens, a dummy variable indicating the order of two policy views questions, and Social Architects' characteristics.

Figure A9 presents the results. We find a negative correlation between the welfare weights assigned to the Recipient in the highest income bracket ($R7$ earning \$500,000) and preferences to increase top-taxes. The effect is significant at the 10% level.[23] We also find a positive correlation between the welfare weights assigned to Recipients 1, 2, 3, and 6 and the preference to increase top-taxes. This finding suggests that inferring people's welfare weights from their policy views can be challenging because we cannot easily uncover the underlying distribution of welfare weights.

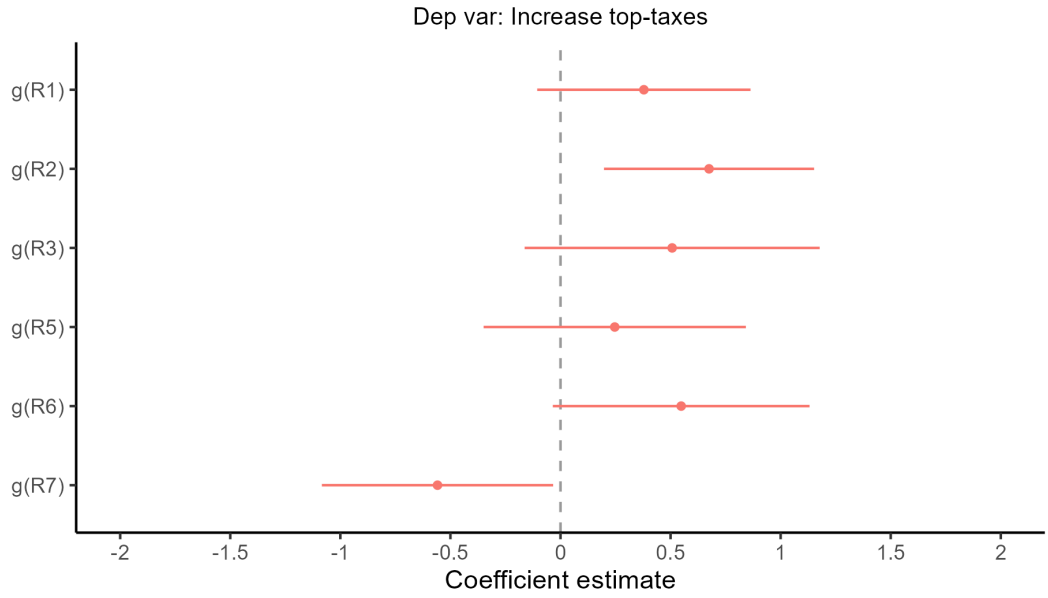


Figure A9: Coefficient Plot of the Regression of Top Taxes on Weights

Notes: The figure presents coefficient estimates. The dependent variable is *Increase top-taxes*, which takes values from 1 to 7, with higher values indicating a greater desire for the government to increase top-taxes. The explanatory variables are the weight assigned to Recipients 1 through 7 (excluding Recipient 4), 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), *Screen Order* (dummy indicating the order of the decisions screens), *Policy Order* (dummy indicating the order of the policy views questions). The regression uses data from Wave 1. Error bars are computed using HC3 standard errors.

[23]The Pearson correlation between $g(R_7)$ and the preference to increase top-taxes is -0.11, which is significant at the 1% level.

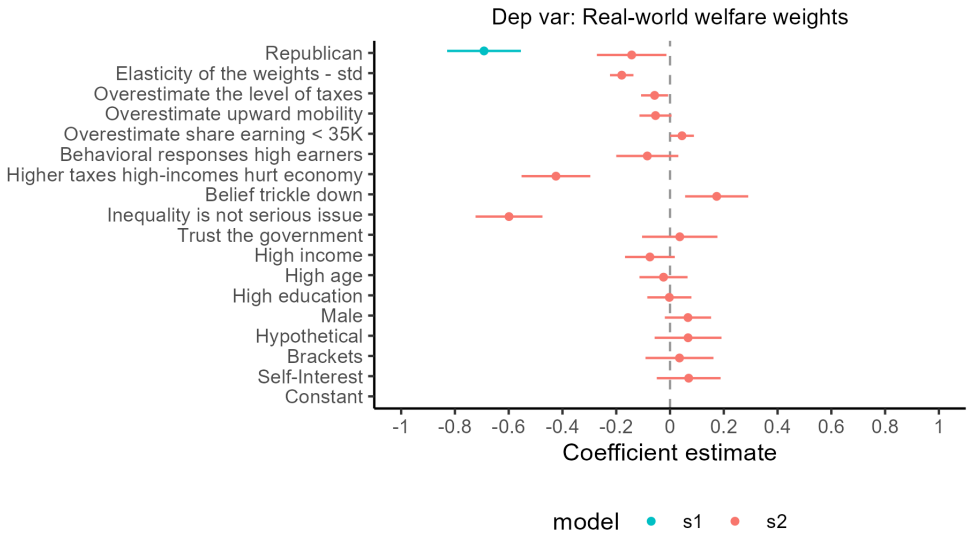
D.6 Decomposing the Partisan Gap in Preferences for Redistribution

Our results in the main text suggest that Social Architects' political affiliation predicts their preferences for redistribution at the margin. In this section, we explore how much of the partisan gap in preferences for redistribution is due to other factors. Our measure of redistribution at the margin is Social Architects' real-world welfare weights, elicited 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 a positive (negative) value indicates that income should be further redistributed by taking from higher-income (lower/ middle-income) individuals and giving to lower/middle-income (higher-income) individuals.

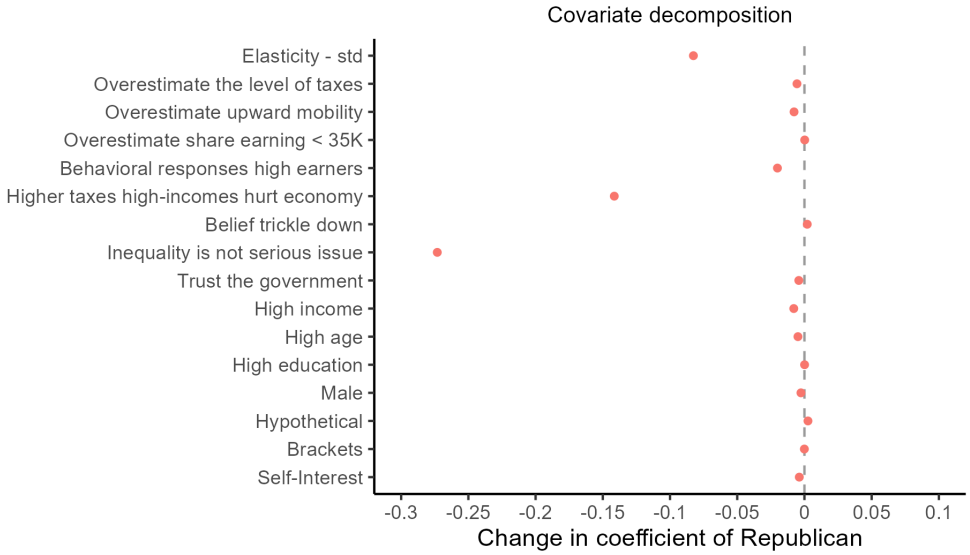
We regress Social Architects' real-world welfare weights (elicited using the survey measure) on Social Architects' political affiliation and various factors. These factors include Social Architects' standardized elasticity of the weights (elicited using the experimental measure), a set of variables capturing Social Architects' misperceptions, and a set of variables capturing Social Architects' views about taxation and government. The standardized elasticity of the weights is the standardized estimated $\hat{v}$ from the regression $log(g(R_j)) = \beta_0 + vlog(\textit{recipient income}_j) + \epsilon_j$, where $g(R_j)$ is the weight assigned by Social Architect i to Recipient j and $\textit{recipient income}_j$ is the disposable income of Recipient j. We standardize the coefficient estimate $\hat{v}$ by subtracting the mean and dividing by the standard deviation. The explanation of the other variables can be found in Appendix Section A.1.

Panel (a) in Figure A10 presents the coefficient estimates of the regression. In Model s1, we only include Social Architects' political affiliation as an explanatory variable, while in Model s2, we include all the variables indicated above as explanatory variables. In both models, Republicans have less progressive real-world welfare weights. 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. This indicates that 79.4% of the partisan gap in policy preferences is driven by the other variables in the model.

Panel (b) in Figure A10 presents the change in the coefficient estimate of *Republican* (between model s2 and s1) that can be attributed to each of the other variables in the regression. The change is computed using a covariate decomposition procedure proposed by Gelbach (2016). We find that about 8% of the partisan gap in preferences for redistribution is driven by welfare preferences. Beliefs about higher taxes on high-income individuals hurting the economy and beliefs about inequality having externalities explain 14% and 27% of the partisan gap.



(a) Regression estimates



(b) Covariate Decomposition

Figure A10: Partisan Gap in Real-world Weights

Notes: Panel (a) presents coefficient estimates. The dependent variable takes values from -2 to +2, with higher values indicating more progressive real-world welfare weights. See the main text for an explanation of the explanatory variables. We reverse code *Inequality is serious issue* in this regression. Models s1 and s2 are weighted using sampling weights. Error bars are computed using HC3 standard errors. Panel (b) presents the change in the coefficient estimate of *Republican* between models s1 and s2 in Panel (a) that can be attributed to the other variables. The change is computed using the covariate decomposition procedure proposed by Gelbach (2016). The figures use data from Wave 2.

E Details on Analyses

E.1 Recipients' Incomes and the Income Distribution

We obtain data on the income distribution from the Distributional National Accounts micro-files of Piketty et al. (2018). For simplicity, we treat all individuals in the data as single filers, regardless of whether they filed taxes or their actual filing status. We partition the data into percentiles (0, 1, . . . ,99, 99.5) based on the distribution of disposable incomes. In particular, we identify the disposable income thresholds that group individuals into percentiles. The disposable income (*diinc*) measure we use includes in-cash redistribution and in-kind transfers. We plot the disposable income thresholds against percentiles, dropping the data on the lowest (0) percentile to make the figure more readable.

E.2 Details on the Calibration of Tax Schedule

We obtain data on the income distribution from the Distributional National Accounts micro-files of Piketty et al. (2018). For simplicity, we treat all individuals in the data as single filers, regardless of whether they actually file taxes and their actual filing status. We use the pre-tax labor income (*plinc*) of individuals, dropping individuals with non-positive labor incomes. We partition the income distribution into percentiles (0, 1, . . ., 99, 99.9, 99.99) and calculate the average income in each percentile.

The tax formula requires a smooth income distribution. To smooth the income distribution, we estimate a smoothing splines regression, regressing the log of the incomes on the percentiles. We set the smoothing parameter as 0.1.[24] The fit from the smoothing spline is used to create a smooth income distribution for 1001 equally spaced percentiles (0 to 100).

[24]We set the spar equal to 0.1. A k-fold cross-validation suggests that a smoothing parameter of less than 0.1 produces the best fit.

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 A11. 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)

$$\begin{aligned} & g_1 C > g_1 A \quad \text{and} \quad g_2 B > g_2 D \\ \Rightarrow & g_1 C - g_1 A > g_2 D - g_2 B \end{aligned} \tag{12}$$

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.

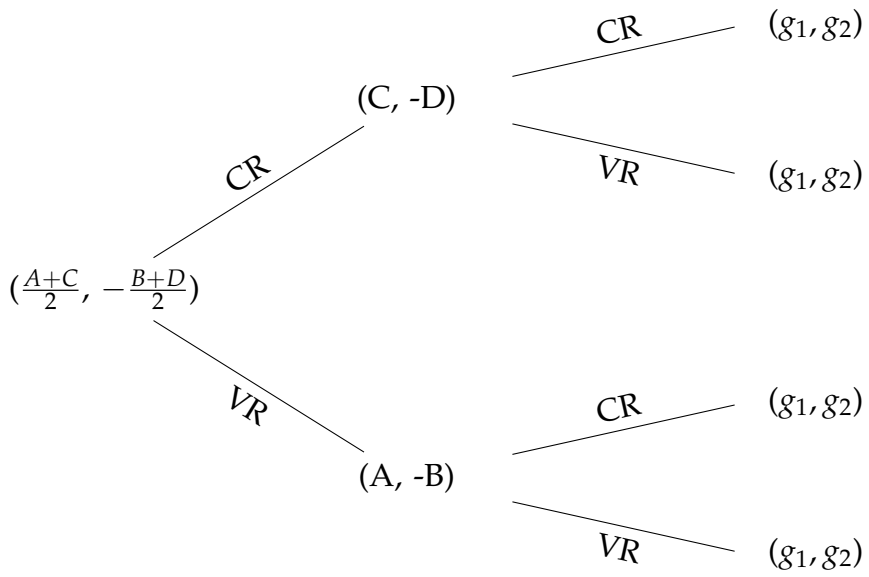


Figure A11: Simple Version of the Staircase Procedure.

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 (12). 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

$$2g_1 500 - 2g_2 500 > g_1(A + C) - g_2(B + D) \tag{13}$$

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 13. 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

$$2g_1 500 - 2g_2 500 < g_1(A + C) - g_2(B + D) \tag{14}$$

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 14. 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

$$g_1 C - g_2 D > g_1 500 - g_2 500 > g_1 A - g_2 B \tag{15}$$

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.

G.1 Treatment Loss x Moderate

[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.

G.2 Treatment Loss x High

[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]

G.3 Treatment Gain x Moderate

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

[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:

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

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

	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 ———————————————

[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 ———————————————

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

[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)]

G.4 Treatment Gain x High

[All screens are identical to the screens in Treatment Gain x Moderate, 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 High.]

H Instructions - Wave 2

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

H.1 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.

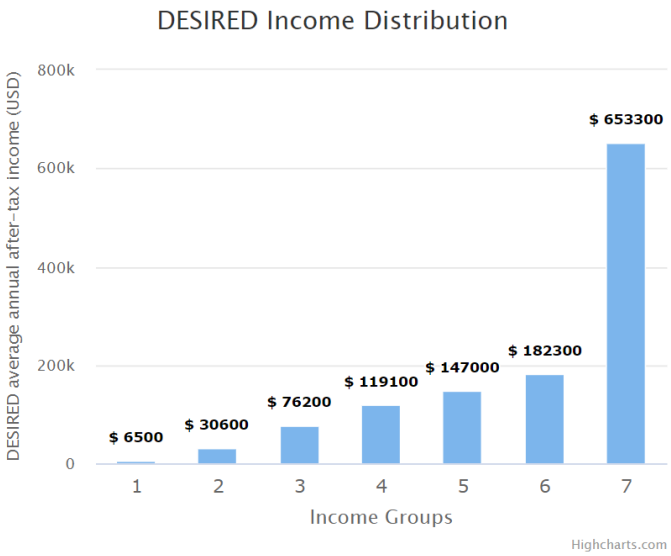
We would like you to indicate what you think the average tax rate for each tax group in the U.S. should be. This can be done as follows. **You can increase or decrease the average tax rates of the first six groups. The average tax rate of group 7 adjusts automatically so that all seven groups together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

H.2 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]

H.3 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]

H.4 Self-Interest Treatment

[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. & Teso, E. (2018), 'Intergenerational mobility and preferences for redistribution', *American Economic Review* **108**(2), 521–554.

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

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

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

Bursztyn, L., Haaland, I., Rao, A. & Roth, C. (2020), Disguising prejudice: Popular rationales as excuses for intolerant expression, Technical Report 27288.

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

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

Chen, D. L., Schonger, M. & 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. & Smeets, P. (2019), 'Why do the rich oppose redistribution? An experiment with america's top 5%', *Working paper* .

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

Erikson, R. S. (2015), 'Income inequality and policy responsiveness', *Annual Review of Political Science* **18**, 11–29.

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

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

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

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

Fong, C. M. & 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.

Haaland, I. & 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. & 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. & 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. & Stantcheva, S. (2015), 'How elastic are preferences for redistribution? evidence from randomized survey experiments', *American Economic Review* **105**(4), 1478–1508.

Lobeck, M. & Støstad Nyborg, M. (2022), 'The consequences of inequality: Beliefs and redistributive preferences', *Working paper* .

Lockwood, B. B. & 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.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. (2022), 'Data quality of platforms and panels for online behavioral research', *Behavior Research Methods* **54**(4), 1643–1662.

Piketty, T. & Saez, E. (2013), Optimal labor income taxation, *in* 'Handbook of public economics', Vol. 5, Elsevier, pp. 391–474.

Piketty, T., Saez, E. & 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. & 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. & Stantcheva, S. (2016), 'Generalized social marginal welfare weights for optimal tax theory', *American Economic Review* **106**(1), 24–45.

Schokkaert, E. & 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. (2021*a*), Understanding economic policies: What do people know and how can they learn?, Technical report.

Stantcheva, S. (2021*b*), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 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. & Jongen, E. L. (2013), 'Optimal redistributive taxes and redistributive preferences in the netherlands', *Working paper* .