Preferences for Taxing Personal Characteristics*

Julien Senn†

Krishna Srinivasan‡

Job Market Paper #1

November 5, 2025

[latest version] [Link to JMP #2]

Abstract

While tagging—the conditioning of taxes on personal characteristics correlated with earning ability—can improve efficiency, it is often assumed to lack political support because it violates horizontal equity, which stipulates treating equals equally. To empirically test this view, we conducted an online vignette experiment with a U.S. general population sample (N = 3,012) and report three results that challenge it. First, support for tagging varies substantially across tags and individuals. Consistent with theoretical prescriptions, it is higher for characteristics with a strong correlation with ability or reflecting needs. However, the immutability of a tag does not predict support, contrary to theory. Second, variation in support reflects both horizontal and vertical equity concerns. Third, other considerations, such as efficiency, also matter but less so. Incorporating fairness can either limit or amplify optimal tagging relative to a canonical utilitarian benchmark. Finally, we compare our results with the tags used in the U.S. tax code and those discussed in the literature.

Keywords: Taxation; Tagging; Inequality; Equity; Fairness

JEL Classification: D63, H21, I31

*We thank Matthew Weinzierl, Roberto Weber, Dmitry Taubinsky, Simona Sartor, Canishk Naik, Ben Lockwood, Emanuel Hansen, Antoine Germaine, Ernst Fehr, Francesco Capozza, Spencer Bastani, Pierre Boyer, Björn Bartling, Sandro Ambühl, and various seminar participants for their helpful comments. We thank Prisha Sharma for her excellent research assistance. This study was pre-registered: https://doi.org/10.1257/rct.14473-1.0

†Sorbonne University: julien.senn@econ.uzh.ch

‡UniDistance: krishna.srinivasan@unidistance.ch

1 Introduction

Income taxation is a primary tool for governments to raise revenue and reduce inequalities. In most countries, individuals' tax liability depends not only on income but also on personal characteristics, such as age, number of children, and disability status (Mankiw et al. 2009). Data collected in this paper shows that the U.S. tax code conditions taxes on several characteristics, including marital status and sector of employment. Determining which characteristics should be used as a basis for taxation is important for tax policy design.

The theoretical literature on optimal taxation argues that taxes should be based on any largely immutable characteristic correlated with earning ability because doing so improves efficiency (Mirrlees 1971, Akerlof 1978).[1] If ability were observed, taxing ability (the "first-best") would not distort choices; since ability is unobserved, income and characteristics serve as proxies. Taxes on personal characteristics, to the extent that they do not depend on choices, are more efficient since they do not distort choices, while income taxes, which depend on labor choices, can distort such choices. Conditioning taxes jointly on income and characteristics—commonly referred to as "tagging"—has been the focus of a growing literature on various characteristics, including height (Mankiw et al. 2009), gender (Cremer et al. 2010), and sector of employment (Gomes et al. 2018).

Despite its theoretical appeal, tagging is commonly assumed to be socially unacceptable—that is, at odds with widely held normative views—because it conflicts with horizontal equity, which stipulates treating "equals" equally (e.g., Fleurbaey & Maniquet 2018, Mankiw & Weinzierl 2010, Saez & Stantcheva 2016). For example, height-based taxation is inappropriate if tall and short individuals are deemed equal. Yet there is little direct evidence on whether citizens view tagging as socially unacceptable. Our paper addresses this gap by conducting an experiment with the U.S. general population ($N = 3012$) that (i) provides descriptive evidence on support for tagging across various characteristics, (ii) examines which types of characteristics people are willing to use, (iii) assesses whether support aligns with theoretical prescriptions, and (iv) analyzes the determinants of support.

To provide descriptive evidence on people's support for conditioning income taxes on various characteristics (or "tags"), we elicit their support for various tags discussed in the literature and those employed by the U.S. tax system. Each tag divides the population into two groups, a "tagged" and an "untagged" group, with the tagged group earning lower

[1] Conditioning taxes on characteristics correlated with labor supply elasticities or preferences can also improve efficiency (e.g., Karabarbounis 2016, Alesina et al. 2011); however, this is not the focus of our paper. Other motivations for conditioning taxes on characteristics include improving redistribution across groups (e.g., Gaubert et al. 2021, Cremer et al. 2003), correcting externalities (e.g., Bovenberg & Jacobs 2005, Lockwood et al. 2017), and encouraging socially desirable behaviors (e.g., Milligan 2005).

incomes on average. Because the tag is correlated with earning ability (proxied by income), optimal tax theory prescribes setting lower taxes for the tagged group relative to the untagged group to increase efficiency. Participants indicate their support for setting lower taxes for the tagged group (henceforth referred to as "receiving a tax break") on a 7-point scale ranging from "+1: Slightly support" to "+3: Strongly support." To make responses consequential, we inform participants that their responses will be shared with U.S. policymakers.

While support for tags is generally low, several tags receive high levels of support. Participants exhibit low support for providing tax breaks for groups such as minorities (16% support this tag), women (19%), and the young (25%); moderate support for tax breaks for individuals living in low-income regions (39%); and high support for providing tax breaks for individuals with dependents (63%), disabilities (60%), and visual impairments (49%). These results are robust to various specifications, including restricting the sample to participants with strong preferences for having their views considered in policymaking and those aware that the tax system already employs tagging, indicating that participants hold meaningful preferences. In addition to observing variation in support across tags, we also observe variation across individuals: while the median participant supports about a third of tags, about 19% support none of the tags, and 3% support all of them. Taken together, the observed variation in support across tags and individuals suggests that attitudes toward tagging are likely shaped by both the characteristics of the tags and individual beliefs and preferences. Furthermore, these findings point to a more nuanced view than accounts in the literature that anticipate finding universally low public support.

Having documented variation in support across tags, we ask which types of tags garner higher support and whether support aligns with the prescriptions from a canonical utilitarian model of tagging. We develop a simple model that yields comparative statics describing how the optimal tax differentiation between the untagged and tagged groups varies with three features of a tag: (i) correlation with ability (strong vs. weak), (ii) mutability (mutable vs. immutable), and (iii) correlation with consumption needs (correlated vs. uncorrelated). The first two are widely studied (Mankiw et al. 2009, Piketty & Saez 2013), while the third is novel and motivated by the literature (e.g., Boadway & Pestieau 2006, Konow 2001).

To assess how these features shape support, we design a vignette experiment that varies these features by treatment. Participants view a scenario that closely mirrors the ones used to elicit support for specific tags; however, the groups defined by the tag are left unspecified. This abstraction allows us to identify the causal effect of relevant features without being confounded by other features that vary across tags. We find that participants are more likely to support tagging when the tag is a strong signal of ability or relates to consumption needs, consistent with theoretical prescriptions. However, the immutability of the tag does

not predict support, contrasting the theoretical prescription of using immutable tags.

To shed light on the concerns underlying support for tagging across individuals and treatments in the vignette experiment, we analyze participants' open-ended explanations using a Large Language Model (LLM). The most prevalent concern among opponents of tagging is horizontal equity, as hypothesized by the literature: participants argue that individuals with the same income should face the same tax burden regardless of group affiliation. Among supporters, vertical equity concerns, conceptualized as reducing inequality across groups, are salient: participants perceive the tagged group as having lower incomes, needs, or disadvantages, and emphasize the need for supporting such groups. These horizontal and vertical equity concerns also help account for the variation across treatments.

While the open-ended analysis hints at the importance of fairness concerns, we also measure such concerns directly. We conceptualize fairness concerns—capturing horizontal and vertical equity—via welfare weights, following the literature (e.g., Berg 2025, Lockwood & Weinzierl 2015, Bergstrom & Dodds 2021). Welfare weights measure society's valuation of an additional dollar of consumption for each group, and are a reduced-form measure of underlying fairness ideals (Saez & Stantcheva 2016). We elicit (qualitative) welfare weights using an incentivized bonus allocation task, following Capozza & Srinivasan (2024). If the tagged and untagged groups have the same average welfare weights, horizontal equity implies no tagging; if their average welfare weights differ, vertical equity supports redistribution from the group with lower to the group with higher welfare weights, thereby rationalizing tagging. We find that fairness concerns—captured by welfare weights—account for the variation in support across individuals and across treatments.

While fairness concerns are a strong predictor of support across individuals and treatments, they do not explain all the variation. We elicit other concerns using closed-ended survey questions and find that efficiency concerns also predict support for tagging across individuals and treatments, though to a lesser extent. Views about the scope of government, implementation concerns, and privacy concerns play a more limited role, helping to explain differences across individuals but contributing little to differences across treatments.

Overall, our results highlight three qualifications to accounts in the literature that anticipate generally low public support for tagging because it violates the fairness principle of horizontal equity: (i) support varies substantially across tags and across individuals, (ii) vertical equity matters alongside horizontal equity, and (iii) concerns beyond fairness also play a role. While the canonical utilitarian model has two goals—improving efficiency and between-group redistribution—our results suggest that participants are likely guided primarily by redistributive motives. Given the importance of such fairness concerns, we incorporate them into our theoretical model to characterize when tagging is socially acceptable.

We model social acceptability via the welfare weights assigned by society to the tagged and untagged groups, which capture their assessment of the value of assigning a $1 transfer to an individual from each group. The optimal tax differentiation between groups is increasing in the difference between their average welfare weights.

Finally, we compare our results on citizens' support for tags with the tags used in the U.S. tax code and those discussed in the literature. We construct a novel database of tags discussed in the literature and those used in the U.S. federal income tax system (coded from U.S. tax law). There is scope for expanding tagging based on existing research, as well as scope for future research to identify tags that are both economically beneficial and socially acceptable. Furthermore, citizens' preferences and existing tax policies are broadly aligned, with a few exceptions.

Our paper contributes to several strands of the literature. First, we add to the empirical literature on measuring people's support for tagging (e.g., Weinzierl 2014, Drenik & Perez-Truglia 2018, Sausgruber & Tyran 2014).[2] Compared to Weinzierl (2014) and Drenik & Perez-Truglia (2018), we examine a broader set of tags, provide causal evidence on preferences for types of tags, and explore underlying mechanisms. Sausgruber & Tyran (2014) study tagging with commodity taxes in the laboratory, whereas we study actual income tax policies.

Second, we contribute to the theoretical literature developing general theories of tagging (e.g., Akerlof 1978, Immonen et al. 1998, Boadway & Pestieau 2006, Mankiw & Weinzierl 2010, Cremer et al. 2010) and the literature focusing on the efficiency benefits of specific tags (e.g., Mankiw & Weinzierl 2010, Gomes et al. 2018), which is summarized in Table 5. We show how horizontal and vertical equity concerns can be incorporated into theory via welfare weights to identify socially acceptable tagging.[3] Related work also aims to incorporate social acceptability: Saez & Stantcheva (2016) incorporate horizontal equity concerns via welfare weights in a Ramsey tax problem, and Weinzierl (2014) incorporates the principle of "equal sacrifice" as a constraint in a model of tagging.

Third, we contribute to the broader literature on conditioning taxes on personal characteristics. One strand, also under the umbrella of tagging, focuses on the efficiency benefits of tagging with characteristics correlated with individuals' elasticity of taxable income (e.g., Alesina et al. 2011, Karabarbounis 2016). Other strands focus on other goals, such as redistributing across groups (e.g., Gaubert et al. 2021, Cremer et al. 2003), correcting externalities

[2] A related strand measures support for targeting policies (e.g., Liscow & Pershing 2022).

[3] Related work conceptualizes horizontal and vertical equity concerns through welfare weights: Berg (2025) decomposes the inverse-optimum welfare weights into horizontal and vertical equity, while other papers model such equity considerations via welfare weights in settings where redistribution addresses ability differences but not differences in preferences (e.g., Fleurbaey & Maniquet 2006, Lockwood & Weinzierl 2015, Bergstrom & Dodds 2021).

(e.g., Bovenberg & Jacobs 2005, Lockwood et al. 2017), and encouraging socially desirable behaviors (e.g., Milligan 2005).

Finally, we contribute to the broader literature of preferences for redistribution (e.g., Cappelen et al. 2007, Saez & Stantcheva 2016, Drenik & Perez-Truglia 2018, Blesse et al. 2019, Kuziemko et al. 2015, Stantcheva 2021, Fehr et al. 2024, Konow 2001, Capozza & Srinivasan 2024). Many of these papers aim to identify which considerations shape preferences for redistribution. A key finding is that people are willing to redistribute based on differences in effort but not differences in preferences (Cappelen et al. 2007, Saez & Stantcheva 2016, Drenik & Perez-Truglia 2018, Blesse et al. 2019). Another finding, consistent with our results, is that people are willing to redistribute based on differences in needs (e.g., Cappelen et al. 2013, Konow 2001). Our paper contributes to this literature by providing a theoretical foundation and a reduced-form empirical measure of horizontal and vertical equity concerns (in the spirit of Capozza & Srinivasan (2024)), which can help account for the observed variation in redistributive preference across contexts.

The remainder of the paper is structured as follows. Section 2 presents the theoretical framework. Section 3 describes the experimental design. Section 4 presents the results. Section 5 synthesizes theory, practice, and citizens' support. Finally, Section 6 concludes.

2 Theoretical Framework

Following Mankiw & Weinzierl (2010), we present a simple theoretical framework that formalizes tagging. We extend this framework to derive comparative statics that characterize how tagging depends on the properties of the characteristic. Finally, we highlight how fairness concerns can be incorporated into the framework.

2.1 Setup

Consider an economy comprising individuals partitioned into two mutually exclusive and exhaustive groups based on an immutable personal characteristic (or *tag*). We denote these groups by $g \in \{t, u\}$, where t denotes the *tagged* group and u denotes the *untagged* group. Individuals in each group are differentiated by their exogenous wages (or *ability*), with I discrete wage levels. Let $\pi_{g,i}$ denote the share of individuals in Group g earning wage w_i, with $\sum_i \pi_{g,i} = 1 \; \forall \; g \in \{t, u\}$. Individuals choose their labor supply $\ell_{g,i}$, earning income $y_{g,i} = w_i \cdot \ell_{g,i}$. Individuals' wages and labor effort are unobserved, while income and group status are observed. Consumption $c_{g,i}$ equals after-tax income, where tax liability varies within groups but not between groups. When the planner can observe group status, the

planner can also use aggregate lump-sum transfers, which vary between groups but not within groups. Denote the aggregate lump-sum transfer paid by Group g as r_g. Individuals' utility is a function of consumption and labor, given by

$$U_{g,i} = u(c_{g,i}, \ell_{g,i}), \tag{1}$$

which is assumed to be increasing and concave in consumption and decreasing and convex in labor effort. The disutility from labor and the elasticity of taxable income are identical across tagged and untagged groups; heterogeneity in preferences or labor-supply elasticities across groups provides an alternative rationale for tagging (e.g., Alesina et al. 2011, Karabarbounis 2016).

2.2 Optimal Taxes

In the standard Mirrleesian framework, the *first-best* solution involves taxing (observable) ability, because this achieves redistribution without distorting choices. Since ability is unobserved, the *second-best* solution involves taxing income—determined by ability and labor effort—which involves balancing the social welfare benefits (redistribution) against the social welfare costs (labor supply distortions).

When the planner can also observe group status, the planner can condition transfers on group status. Formally, the social planner's objective is to choose income and consumption bundles to maximize a utilitarian social welfare function that is uniform and linear in individuals' utilities (formalized in Appendix Section A).[4] The planner is subject to two constraints: (i) ability and effort are unobserved (though income and group status are observed), and (ii) taxes are purely redistributive, requiring the budget to be balanced. A well-known result in the literature is that the optimal consumption bundles must satisfy

$$\sum_{i=1}^{I} \pi_{t,i}\, u_c(c_{t,i}, \ell_{t,i}) = \sum_{i=1}^{I} \pi_{u,i}\, u_c(c_{u,i}, \ell_{u,i}) \tag{2}$$

where $u_c(c_{g,i}, \ell_{g,i})$ is the marginal utility of consumption of an individual with wage w_i in Group g. Condition (2) states that, at the optimum, the average marginal utility of consumption is equalized across groups. Intuitively, if the averages differed, social welfare can be increased by transferring \$1 from the group with the lower to the group with the higher marginal utility of consumption. The proof can be found in Appendix Section A.

[4]The results extend to any social welfare function that is concave in individuals' utilities (Mankiw & Weinzierl 2010).

When the tagged group has a higher average ability and, consequently, a higher average marginal utility of consumption, a lump-sum transfer is paid from the untagged to the tagged group. With the budget constraint on the transfers $r_t + r_u = 0$, the between-group transfer paid to the tagged group is given by $r = -r_t = r_u$. Because this transfer depends only on group status, it improves equity (redistribution across ability types) without decreasing efficiency. The following proposition formalizes the intuition that tagging emerges as part of the optimal policy.

Proposition 1 *When an observable and immutable personal characteristic is correlated with earning ability, it is optimal to condition taxes on that characteristic.*

The preceding discussion shows that tagging achieves two objectives. First, it improves efficiency by reducing the reliance on distortionary income taxes and instead relying on non-distortionary between-group transfers. Second, it reduces the between-group consumption inequality using between-group transfers.

2.3 Comparative Statics

We derive three comparative statics that describe how the optimal between-group transfer (r) varies with three properties of tags: (i) immutability, (ii) strength of the correlation with ability, and (iii) correlation with needs. The first two relate to efficiency and have been widely studied (Mankiw et al. 2009, Piketty & Saez 2013), while the third relates to welfare and is motivated by the discussion of needs in the taxation literature (Piketty & Saez 2013, Boadway & Pestieau 2006, Konow 2001). Proofs of the propositions can be found in Appendix Section A. These results have appeared in various forms in Parsons (1996), Immonen et al. (1998), Viard (2001*a*,*b*), and Salanié (2002).

First, consider the strength of the correlation between the tag and ability, defined as the difference in average ability across groups. As this correlation strengthens, the tagged group's average ability falls further below that of the untagged group, widening the gap in their incomes and (hence) consumptions. Since marginal utility is strictly decreasing in consumption, the tagged group's average marginal utility rises relative to the untagged group's. To restore the optimality condition in Equation (2), the planner increases the between-group transfer to the tagged group. We summarize this in Proposition 2.

Proposition 2 *The optimal between-group transfer is increasing in the strength of the correlation of the tag with ability.*

Second, we compare immutable tags with mutable tags. With immutable tags, group membership is fixed; with mutable tags, some individuals in the untagged group may misreport their group status to benefit from the lower taxes afforded to those in the tagged group. For example, if taxes depend on disability status, non-disabled individuals might falsely claim to be disabled to take advantage of reduced tax rates. Such switching weakens the tag's information about underlying ability, which induces the planner to reduce the size of the between-group transfer.[5] This yields Proposition 3.

Proposition 3 *The optimal between-group transfer is higher when the tag is immutable than when it is mutable.*

Third, consider differences in consumption needs. Following Boadway & Pestieau (2006), define needs-adjusted consumption as $\tilde{c}_{g,i} = y_{g,i} - t_{g,i} - n_{g,i}$, where $n_{g,i}$ denotes consumption needs.[6] Individuals with higher consumption needs require more resources to attain the same utility as others. We assume that needs are not observable at the individual level, but the distribution of needs of the two groups is observed. We measure the strength of the correlation of the tag with needs by the difference in average needs across groups. As this correlation strengthens, the tagged group's average needs-adjusted consumption falls relative to the untagged group's, raising its average marginal utility of consumption. To satisfy Equation (2), the planner increases the between-group transfer to the tagged group. This is formally stated in Proposition 4.

Proposition 4 *The optimal between-group transfer is increasing in the strength of the correlation of the tag with consumption needs.*

Propositions 2 and 4 both address disparities in consumption across groups. We state them separately because citizens may distinguish ability-driven from needs-driven differences in consumption when they know the distribution of needs across groups.

2.4 Incorporating Fairness Concerns

We now incorporate fairness concerns into the framework to identify when tagging is socially acceptable. A common view is that citizens may oppose tagging because it violates *horizontal equity*—the principle that "equals" should be treated equally (e.g., Fleurbaey & Maniquet

[5] If switching occurs due to behavioral responses other than evasion, such as individuals intentionally incurring health risks to be classified as disabled, there are additional efficiency costs that arise from such choice distortions, which further lower the appeal of mutable tags.

[6] See Pestieau & Racionero (2015) for tagging with *leisure needs*—needs increase with ability.

2018, Saez & Stantcheva 2016, Mankiw et al. 2009). In our context, horizontal equity implies that if the tagged and untagged groups are perceived as equals, they should face the same tax treatment. However, horizontal equity alone is not sufficient: it does not explain why we observe differential treatment of groups in practice (see evidence in Section 5) and citizens supporting some tags (see evidence in Section 4.1). A concern for reducing inequality across groups deemed different is captured by *vertical equity*. Thus, horizontal equity requires no tagging for groups judged as equals, while vertical equity provides a justification for tagging when the groups are deemed different. To incorporate such fairness concerns, we introduce *Pareto weights* $\omega_{g,i}$, which capture the planner's normative priority on the utility of an individual with wage w_i in Group g. With Pareto weights, Equation (2) changes to

$$\sum_{i=1}^{I} \pi_{t,i}\, \omega_{t,i}\, u'(c_{t,i}) \;=\; \sum_{i=1}^{I} \pi_{u,i}\, \omega_{u,i}\, u'(c_{u,i}), \tag{3}$$

where the term $\omega_{g,i}\, u'(c_{g,i})$ is the *social marginal welfare weight* (henceforth welfare weights) for an individual in group g with wage w_i. Welfare weights measure society's valuation of providing an additional dollar of consumption to different individuals. They are a reduced-form representation of underlying fairness ideals (Saez & Stantcheva 2016).

Welfare weights provide a unified way to represent both horizontal and vertical equity.[7] In the canonical benchmark model, the Pareto weights are uniform across groups—redistribution between groups is driven solely by differences in their average marginal utility of consumption. There are four deviations from this benchmark achieved by modifying the Pareto weights. First, assigning a higher average Pareto weight to the tagged group strengthens the case for tagging relative to the benchmark: the planner increases the between-group transfer toward the tagged group beyond what consumption differences alone would imply. Second, assigning a higher average Pareto weight to the untagged group weakens the case for tagging relative to the benchmark. Third, if Pareto weights are chosen to equalize the welfare weights across groups, horizontal equity requires no between-group transfers. Fourth, if the untagged group's average Pareto weight is sufficiently high, the direction of the between-group transfer reverses: the transfer flows to the untagged group, even though it has a higher average consumption.

[7](for related approaches, see Berg 2025, Lockwood & Weinzierl 2015, Bergstrom & Dodds 2021). Modeling both horizontal and vertical equity with welfare weights is consistent with work arguing that horizontal equity lacks an independent normative basis (e.g., Kaplow 1989, Galbiati & Vertova 2006, Berg 2024).

3 Survey Design and Data

To provide descriptive evidence on citizens' support for tagging based on various tags, we elicit participants' support for policy proposals involving tagging with various tags (Section 3.1). Next, to understand which types of tags garner higher support, we conduct a vignette experiment (Section 3.2). Finally, we include several questions to understand which concerns drive citizens' support for tagging (Section 3.3). Appendix Figure A1 provides an overview of the survey. The full set of instructions can be found in Appendix Section F. We pre-registered the experimental design and adhered to this plan without deviations.[8]

3.1 Support for Tags

We elicit participants' support for tagging based on various tags. Each tag divides the population into two mutually exclusive groups: a "tagged" and an "untagged" group.[9] Following the theoretical setup, we provide participants with information about the incomes of the groups, which is a proxy for their underlying ability. In particular, we inform participants that the tagged group has a lower average pre-tax income than the untagged group, and the income distributions of the groups overlap. Figure 1 presents a screenshot of one of the tags presented to participants, in which the tagged group includes individuals with visual impairments and the untagged group includes all others. We omit the exact incomes to reduce the cognitive burden on participants.

For each tag, participants are presented with a policy proposal in which individuals in the tagged group would owe slightly lower income taxes than those in the untagged group.[10] We highlight the horizontal distributional implications of the policy by informing participants that, for any given income, an individual in the tagged group would owe lower taxes than an individual in the untagged group. Under optimal tax theory, such a policy can increase efficiency (see Proposition 1). Participants indicate their support for the policy on a 7-point scale ranging from "-3: Strongly oppose" to "+3: Strongly support."

Participants evaluate ten policy proposals, each corresponding to a different tag. We include tags discussed in the literature, as well as those employed by the federal income tax system. When a tag appears in the literature, we define the tagged group based on the

[8] See https://doi.org/10.1257/rct.14473-1.0

[9] Many tags naturally divide the population into two groups (e.g., visual impairment status). For tags that can be divided into multiple groups (e.g., age), we split them into two groups for simplicity.

[10] We use a "gain" frame that focuses on the gains to the tagged group rather than a "loss" frame that focuses on the losses of the untagged group. If participants are loss-averse, the loss frame would garner lower support. Thus, levels of support from a gain frame can be interpreted as an upper bound. However, the difference in fairness perceptions between the two frames is likely to be small (Capozza & Srinivasan 2024), particularly in domains where participants do not make decisions for themselves (Andersson et al. 2016).

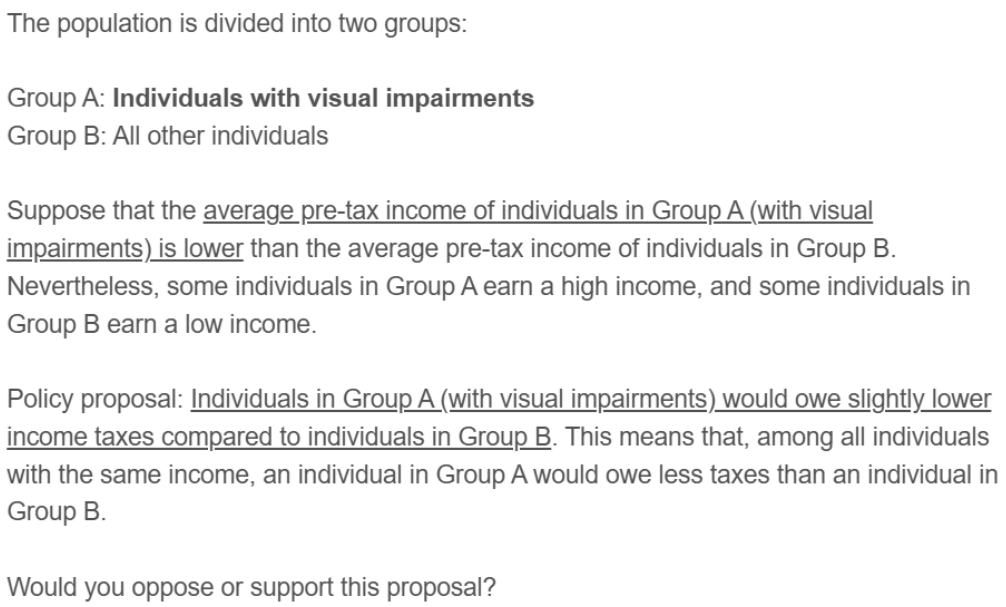


Figure 1: Screenshot of a Policy Proposal

groups in the numerical simulations; when it is not, we instead rely on the groups that receive tax breaks in practice. Table 5 indicates the selected tags and groups, which include: (1) dependents (presence of dependents), (2) disability status (disabled), (3) visual impairment status (visually impaired), (4) region of residence (residing in the bottom 1% of regions by average income and poverty rate), (5) age (individuals aged 18 to 30), (6) marital status (not married), (7) level of education (highest education is a high school diploma or less), (8) gender (women), (9) race and ethnicity (ethnic/racial minorities), and (10) sector of employment (not service sector).[11] The order of the policies is randomized across participants.

The average income differences between the tagged and untagged groups vary across tags (see Appendix Figure A3). The largest difference is by region: individuals in low-income regions earn $82,033 less on average than those in other regions. The second-largest difference is by education: low-educated individuals earn about $49,073 less on average than others. Differences for all remaining tags range from roughly $19,000 to $31,000.

[11] Among the tags documented in the literature, we exclude genes and accidental bequests because they are difficult to define in policy-relevant terms. We exclude height (an immutable tag) to maintain a balance between immutable and mutable tags. We exclude presence of children because, in the U.S., individuals with children earn higher incomes on average. Among the tags used in practice, we drop homeownership status, veteran status, and self-employment status because these groups earn higher incomes. Although veterans and the self-employed have lower labor incomes, their total incomes—the measure of income described in the instructions—are higher. We exclude students because they have very little labor income.

3.2 Vignette Experiment

To understand which types of tags garner higher support, we conduct a vignette experiment. The experimental control allows us to isolate features common across tags that influence support, while minimizing the influence of idiosyncratic features. We focus on features that affect the optimal level of between-group transfers (see Section 2), which allows us to test whether people's support for tagging aligns with the theoretical prescriptions.

In the vignette experiment, participants view a scenario that closely mirrors the ones used to elicit support for specific tags (Section 3.1); however, the groups defined by the tag are left unspecified and labeled merely as Group A and Group B. This abstraction enables us to minimize the influence of idiosyncratic features of tags. Appendix Figure A2 presents a screenshot of the instructions in one of the scenarios. We elicit participants' support for a policy proposal in which individuals in the tagged group (Group A) would owe slightly lower taxes than those in the untagged group (Group B). Theory would prescribe such a policy (see Proposition 1). Their responses can range, on a 7-point scale, from "-3: Strongly oppose" to "+3: Strongly support."

Treatments

Participants are randomly assigned to one of eight treatments in a $2 \times 2 \times 2$ between-subjects design. These treatments consist of variations of the scenario described above. Each treatment dimension considers a feature of tags that affects the optimal level of between-group transfers in theory (see Section 2). Table 1 provides an overview of the treatments.

The first treatment dimension varies the strength of the correlation between the tag and individuals' ability—conceptualized as the overlap in the ability distributions across the tagged and untagged groups. Because ability is unobserved, we implement this by varying the overlap of the groups' income distributions. In Treatment Low-Corr, participants learn that "many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income." In Treatment High-Corr, participants learn that "very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income." The tag has a stronger correlation with ability in Treatment High-Corr because of the lower overlap in the income distributions of the groups.

The second treatment dimension varies whether the tag distinguishing the two groups is immutable or mutable. In Treatment Immutable, participants learn that individuals "cannot influence the group to which they belong," while in Treatment Mutable, they learn that "individuals might be able to influence the group to which they belong."

The third treatment dimension varies whether the tag is correlated with individuals'

consumption needs, measured by the expenditure required to meet a basic needs benchmark. In the Uncorr-Needs treatment, "individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs." In the Corr-Needs treatment, "individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs." Thus, the tag is correlated with needs in Treatment Corr-Needs and uncorrelated with needs in Treatment Uncorr-Needs. Participants are presented with a description of basic needs adapted from Liscow & Pershing (2022).

Table 1: Overview of Treatments

	Correlation with ability			
	Weak		Strong	
	Correlated with needs?			
Mutable?	No	Yes	No	Yes
No	Immutable × Low-Corr × Uncorr-Needs	Immutable × Low-Corr × Corr-Needs	Immutable × High-Corr × Uncorr-Needs	Immutable × High-Corr × Corr-Needs
Yes	Mutable × Low-Corr × Uncorr-Needs	Mutable × Low-Corr × Corr-Needs	Mutable × High-Corr × Uncorr-Needs	Mutable × High-Corr × Corr-Needs

Quality of Responses

We take several steps to obtain high-quality responses. First, because the vignette is abstract, we place it after the policy proposals involving various tags so that participants encounter multiple tags before stating their preferences in the vignette. Second, to make participants' responses consequential, we inform them that the study's findings will be shared with U.S. policymakers and made publicly available, following the approach of Elias et al. (2019). In addition, we include an incentivized redistribution task and test whether it correlates with support for tagging in the vignette (details in Section 3.3). Third, to reduce the role of social desirability bias, we inform participants that there are no right or wrong answers, emphasizing that we want their honest opinions. Fourth, to motivate attentive responses and ensure comprehension, we included one attention check and three comprehension checks in the survey, excluding individuals who failed one or more checks. The three true-or-false comprehension questions are designed to test participants' understanding of the key information distinguishing the treatments. For example, in the Uncorr-Needs treatment, participants are asked whether the following statement is true or false: "Individuals in

Group A and Group B spend the same amount of money, on average, to meet their basic needs."

We also examine support among those with high-quality responses, including two questions that serve as proxies for response quality. First, we assessed participants' knowledge about tagging by asking them whether the U.S. tax system conditions taxes on personal characteristics. Second, we asked participants to express the extent to which public authorities should consider their support for tags when deciding on tax policies. We reasoned that participants who believe their responses may be of low quality, because of their unfamiliarity with tagging or because of "trembling-hand" mistakes, would have stronger preferences to exclude their views from consideration.

3.3 Concerns Explaining Support for Tagging

We employ three complementary strategies to uncover participants' concerns regarding tagging. First, to capture participants' "first-order" concerns (Haaland et al. 2025), we include an open-ended question immediately after the vignette experiment that asks participants to explain their reasoning for their support.

Second, we capture fairness concerns, conceptualized as welfare weights (see Section 2). We measure welfare weights using an incentivized bonus allocation task, following Capozza & Srinivasan (2024). After the vignette experiment, participants choose whether to allocate a \$500 bonus to (i) an individual from the tagged group, (ii) an individual from the untagged group, or (iii) either individual (indicating indifference). Participants' decisions are incentivized: the choice of one randomly selected participant in the study will be implemented.[12] Allocating the bonus to the tagged group reflects higher average welfare weights to the tagged group, which we refer to as progressive weights as the tagged group has a lower average income. Similarly, allocating the bonus to the untagged group reflects regressive welfare weights and indifference implies an equal weight to the groups.[13]

Third, to capture overarching concerns, we include several closed-ended items that measure specific concerns highlighted in the literature (e.g., Stantcheva 2021, Blesse et al. 2019, Slemrod 2025), grouped into five categories: equity, efficiency, scope of government, implementation, and privacy. Appendix Section B provides the definitions of the variables constructed from these questions.

[12]We will recruit the recipient of the bonus such that their characteristics match those described in the vignette experiment.

[13]If a participant assigns average welfare weights g_T to the tagged group and g_N to the untagged group, they value a payment of \$500 to the tagged group at $g_T \times 500$ and to the untagged group at $g_N \times 500$. Allocating the bonus to the tagged group implies $g_T \times 500 > g_N \times 500$, or $g_T > g_N$. Similarly, allocating the bonus to the untagged group implies $g_T < g_N$, and indifference implies $g_T = g_N$.

3.4 Background Characteristics

We collected a rich set of background characteristics to compare the sample to the U.S. population and to explore support for tagging among various sub-groups. These include state of residence, birth year, gender, education, household income, region of residence (zip code), marital status, number of dependents, ethnicity (whether Hispanic), race, employment status, sector of employment, visual impairment status, disability status, and political affiliation.

3.5 Implementation and Sample

We recruited participants from the U.S. general population using the data service provider Bilendi, which is commonly used in the literature (e.g., Stantcheva 2021, Jäger et al. 2024). Participants were rewarded by Bilendi using various methods, such as cash and points in reward programs. At the start of the survey, participants provide their age, gender, education, state of residence, and income. We defined demographic quotas based on these characteristics and excluded participants when the quotas were full. We included an attention check and three comprehension checks, dropping participants failing one or more checks. The data collection began on October 8, 2024, and lasted three weeks. After dropping participants based on our pre-registered criteria, our final sample includes 3,012 participants.[14] The median response time for the survey was 15 minutes.

Table 2 presents the average characteristics of our sample and the U.S. population. Since we included demographic quotas, our sample closely matches the population. However, there are small differences (less than 6 percentage points) for some characteristics because we relaxed the quotas towards the end of the data collection. As an additional check of representativeness, we compare responses in our sample to a question on trust in government with those from Pew Research Center, which used a similarly worded question.[15] In our sample, 19% trust the government "most of the time" or "just about always," closely matching the share (22%) reported in Pew's national poll in 2023. The sample characteristics are balanced across the eight treatments, except for those aged 55 to 65 (Appendix Table A2).

[14] We recruited 8296 participants. We dropped participants who did not consent to participate in the study (2.3%), who arrived after the demographic quotas were full (23.8%), who failed the attention check (6.5%), and who failed the comprehension checks (40.1%). The shares failing the first comprehension question (17.8%), the second question (18.8%), and the third question (19.4%), are not statistically significantly different from each other ($F = 2.17, p = 0.12$).

[15] See https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/

Table 2: Summary Statistics

Variable	Sample	Population
Female	0.55	0.51
Income: $<$ 30,000	0.18	0.13
Income: 30–59,999	0.23	0.18
Income: 60–99,999	0.22	0.23
Income: 100–149,999	0.17	0.20
Income: $\geq$ 150,000	0.20	0.26
Education: Up to Highschool	0.34	0.37
Education: Some college	0.18	0.20
Education: Bachelor or Associate	0.33	0.30
Education: Masters or above	0.15	0.13
Age: 18–34	0.13	0.11
Age: 25–34	0.12	0.17
Age: 35–44	0.16	0.17
Age: 45–54	0.17	0.16
Age: 55–64	0.18	0.16
Age: 65+	0.25	0.23
Region: Northeast	0.19	0.17
Region: Midwest	0.20	0.20
Region: South	0.34	0.39
Region: West	0.27	0.24
Republican	0.30	0.27

Notes: The table presents the average background characteristics of our sample and the U.S. population. The U.S. population demographics are computed using the 2023 American Community Survey (ACS) 1-year estimates for individuals aged 18 and older. The population share of Republicans is the average share of individuals identifying as Republican, based on multiple surveys conducted by Gallup in 2023.

4 Results

4.1 Descriptive Evidence on Support for Various Tags

We begin by presenting descriptive evidence on participants' support for tagging based on various tags. Each tag identifies a tagged group that would owe lower taxes (henceforth referred to as "receiving a tax break"). A participant is defined as supporting a tag if their responses range from "+1: Slightly support" to "+3: Strongly support."

Variation in support across tags: Support for tags is generally low, but several tags receive high levels of support (see Figure 2). Participants exhibit low support for tax breaks

for those not employed in the service sector (15%), minorities (16%), women (19%), those with low education (20%), the unmarried (22%), and the young (25%); moderate support for tax breaks for individuals living in low-income regions (39%); and high support for providing tax breaks for individuals with dependents (63%), disabilities (60%), and visual impairments (49%). Averaging across all tags, 33% of participants support tagging. Support levels are very similar when applying sampling weights to match the population (Appendix Figure A4). Furthermore, the ranking of tags is very similar when support is measured using a continuous scale rather than a binary indicator (Appendix Figure A5).

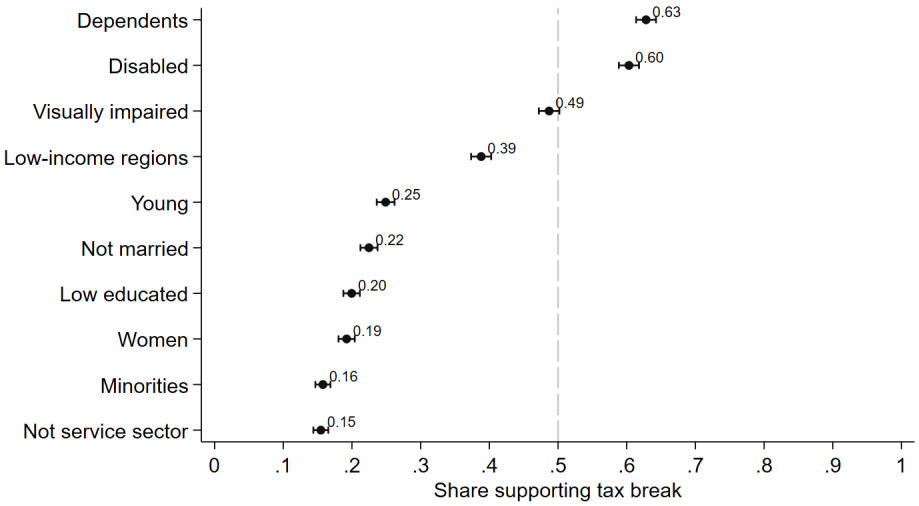


Figure 2: Share Supporting Tags

Notes: The figure presents the share of participants who support providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the share of participants supporting a tax break for each group. A participant is defined as supporting a tax break if their responses range from "+1: Slightly support" to "+3: Strongly support." The bars represent the 95% confidence intervals around the means.

Robustness to Quality of Responses and Self-Interest Motives: We test whether participants' support is robust to low-quality responses. We find that support for tags is very similar among the 38% of participants who express strong preferences for their responses to be considered by policymakers and among the 74% of participants who are aware that the tax system already employs tagging (Appendix Figures A6 and A7). Reassuringly, the high share of participants (74%) aware that the tax system already employs tagging suggests that participants are familiar with the context described in the experiment. Next, we test whether participants' support is robust to self-interest motives. Following the tradition in welfare economics, support for policies unconfounded by self-interest motives provides a normatively appealing policy prescription. We find that support for tags is very similar when

we exclude would-be beneficiaries for each tag—those whose support is prone to self-interest motives (Appendix Figure A8).[16]

Variation in support across individuals: There is also considerable variation in support across participants (Appendix Figure A9). About 19% of the participants support none of the tags, while 3% support all. The median participant supports about a third of tags. There is also heterogeneity across participants within each tag (Appendix Figure A10).

Correlation structure: We summarize the observed variation in support using a principal components analysis (PCA), which identifies components that are weighted averages of participants' support for tags (details in Appendix Section E.1). We identify two components that explain 63% of the variation in the data. The first component, which accounts for about half of the overall variation in the data, assigns positive weights to all tags—it captures a general tendency to be more or less supportive of tagging overall. Consistent with this result, we observe that all pairwise correlations in support for tags are positive (Appendix Figure A11). The second component, which explains 13% of the variation, assigns positive weights to tags with high support (including dependents, disabled, and visually impaired) and negative weights to tags with low support (such as young, not married, or not in the service sector)—it reflects the contrast between popular and unpopular tags.

Summary and Implications: Overall, we find considerable variation in participants' support for tagging both across tags and across individuals, with support being robust to low-quality responses and self-interest motives. A principal components analysis shows that most of the variation in the data can be summarized by two broad dimensions: an overall disposition toward tagging and a contrast between tags with high and low support. Taken together, these results suggest that attitudes toward tagging are shaped by both the characteristics of the tags and individual beliefs and preferences. The next few sections aim to uncover these determinants. Furthermore, these findings point to a more nuanced view than accounts in the literature that anticipate generally low public support for tagging because it violates horizontal equity.

[16]Excluding would-be beneficiaries reduces support by 5–9 percentage points for providing a tax break to the young, those not married, and minorities; because these tags already have low baseline support, our policy prescriptions are unchanged. These results are driven by the low share of would-be beneficiaries for several tags and the limited role of self-interest among these beneficiaries (Appendix Table A13). For several tags, would-be beneficiaries are somewhat more likely to support that tag, but differences are modest or absent for many tags. The largest difference is for age: support for a tax break to the young is 26 percentage points higher among the young than among other participants.

4.2 Which Types of Tags Garner Higher Support?

Our results in the previous section show that there is considerable variation in support across tags. In this section, we analyze the results from the vignette experiment to understand which types of tags garner higher support and test whether participants' support aligns with the prescriptions from optimal tax theory. Figure 3 presents the share of participants who support tagging in the vignette experiment by treatment. In each treatment dimension, the bars average over the treatments in the other two dimensions. Table 3 presents the corresponding regressions, where the dependent variable equals 1 if a participant supports tagging and 0 otherwise, and the explanatory variables are treatment indicators.

First, we compare Treatments Low-Corr and High-Corr. In Treatment Low-Corr, where the tag has a weak correlation with ability, 26% of participants support tagging. In Treatment High-Corr, where the correlation is strong, the share of participants supporting tagging increases to 49%. This 22.7 percentage points difference is statistically significant (Table 3, Column 1). With a baseline support of 26%, the treatment effect corresponds to an 87% increase in support, which is a large effect. This result aligns with Proposition 2, which shows that the optimal between-group transfer increases with the strength of the correlation between the tag and ability.

Second, we compare Treatments Immutable and Mutable. In Treatment Immutable, the individuals described in the vignette cannot influence the group to which they belong, whereas in Treatment Mutable, individuals might be able to influence their group status. The share of participants supporting tagging is very similar between the two treatments: 38% in Treatment Immutable versus 37% in Treatment Mutable. The 1.1 percentage points difference in support across treatments is not statistically significant (Table 3, Column 2). This result contrasts with Proposition 3, which shows that the optimal between-group transfer is higher for immutable than for mutable tags.

Finally, we compare Treatments Uncorr-Needs and Corr-Needs. In Treatment Uncorr-Needs, where the tag is uncorrelated with needs, 31% of participants support tagging. In Treatment Corr-Needs, where the tag is correlated with needs, support rises to 44%. The 12.4 percentage points difference between the treatments is statistically significant (Table 3, Column 3). This finding aligns with Proposition 4, which shows that the optimal between-group transfer is transfer increases with the strength of the correlation between the tag and consumption needs.

While the vignette experiment is designed to study comparative statics, we briefly describe the patterns in the level of support. Support ranges widely across treatments—from 16% when the tag is a weak signal of ability, unrelated to needs, and mutable, to 56% when

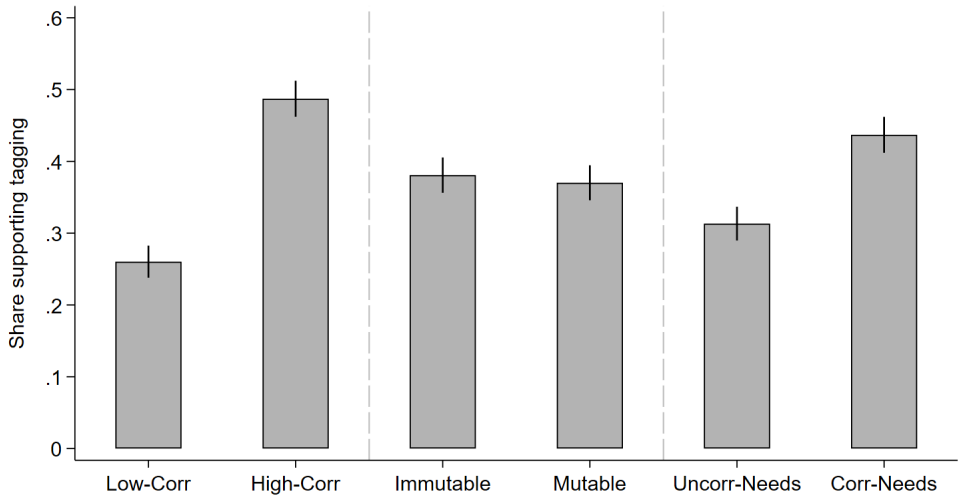


Figure 3: Share Supporting Tagging by Treatments

Notes: The figure presents the share of participants supporting tagging in the vignette experiment by treatment. A participant is defined as supporting tagging if their responses range from "+1: Slightly support" to "+3: Strongly support." Vertical bars represent the 95% confidence intervals around the means.

Table 3: Share Supporting Tagging by Treatments

	(1)	(2)	(3)	(4)	(5)
High-Corr	0.227***			0.228***	0.227***
	(0.017)			(0.017)	(0.017)
Mutable		-0.011		-0.012	-0.018
		(0.018)		(0.017)	(0.017)
Corr-Needs			0.124***	0.126***	0.128***
			(0.018)	(0.017)	(0.017)
Constant	0.260***	0.381***	0.313***	0.203***	0.303***
	(0.011)	(0.013)	(0.012)	(0.016)	(0.025)
Observations	3012	3012	3012	3012	2983
Controls?	No	No	No	No	Yes

Notes: The table reports coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *High-Corr*, *Mutable*, and *Corr-Needs* are treatment indicators. Controls include *Republican*, *High Age*, *High Income*, *High Education*, and *Female*. *Republican* is an indicator for Republicans. *High Age* takes a value of 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. HC3 standard errors in parentheses. * p <0.10, ** p <0.05, *** p <0.01

it is a strong signal of ability, correlated with needs, and immutable (Appendix Figure A12).

Robustness: Our results are robust across several alternative specifications. The treatment effects are similar in a multivariate specification that includes all three treatment indicators jointly (Table 3, Column 4), and when we add background characteristics as controls to this multivariate specification (Table 3, Column 5). Furthermore, the significance levels are intact when we consider continuous measures of support, though the effect sizes of the two significant coefficients move closer together (Appendix Table A3). The treatment effects hold across the full distribution of support (Appendix Figure A13). Each treatment effect within a given dimension is stable across all treatments of the other two dimensions (Appendix Table A4). Finally, the treatment effects are stable across demographic subgroups (Appendix Tables A5 and A6); we only find weak evidence of heterogeneity by partisanship at the 10% level, with smaller treatment effects for Republicans relative to others.

Validation: We present three validation exercises: the first two compare support for tagging in the vignette experiment (based on a controlled scenario) and support for various tags (based on realistic scenarios) at the aggregate and individual levels, and the third tests whether the observed individual-level correlation is driven by the salience of specific tags. First, at the aggregate level, the share of participants supporting tagging in the vignette (38%, pooling across treatments) closely matches the share supporting the individual tags (33%, averaging across tags). Second, at the individual level, participants who support tagging in the vignette experiment support 49% of the tags, whereas those who oppose tagging in the experiment support only 23% of the tags—a 26 percentage points difference that is statistically significant (t-test, $p < 0.01$). Third, we find that the correlation of support in the vignette with support for individual tags ranges from 0.34 to 0.5, while the correlation with average support across tags is 0.57. Thus, support for tagging in the vignette likely captures general dispositions toward tagging rather than the salience of any specific tag.

Summary: Overall, the results from the vignette experiment suggest that participants are more likely to support tagging when the tag is a strong signal of ability or is related to consumption needs, consistent with theoretical prescriptions. Whether a tag is mutable appears to have little impact on support, contrasting the theoretical prescription of using immutable tags. The following sections aim to uncover the underlying concerns driving these treatment effects.

4.3 Underlying Concerns: First-Order Concerns

To uncover participants' underlying concerns regarding tagging, we begin by examining their first-order concerns regarding tagging, drawing on their open-ended explanations of their indicated support in the vignette experiment. We analyze these responses using OpenAI's Large Language Model (LLM). We first compare responses between supporters and opponents of tagging, and then compare responses across treatments. Details and model outputs are provided in Appendix Section E.2.

Support vs. No Support: Among participants who oppose tagging, about 70% emphasize horizontal equity: individuals with the same income should face the same tax burden regardless of group affiliation. Efficiency considerations also emerge, with about 11% of participants citing the complexity of the system and the potential for manipulation. Other themes include the desire for more information, worries that group-based taxation could create unnecessary divisions and conflicts, and neutral or indifferent views. By contrast, participants who support tagging tend to emphasize vertical equity, defined as inequality reduction across groups: 40–45% highlight the need to support lower-income individuals, 27% stress the importance of helping those with higher basic needs expenses, and 25% argue in favor of assisting disadvantaged groups.

Treatment Comparisons: Next, we compare responses across treatments. In the Low-Corr and High-Corr treatments, horizontal equity emerges as the most prominent concern, cited by about 70% of participants in the Low-Corr treatment and 62% in the High-Corr treatment. Other topics appear in roughly similar proportions across the two treatments. The slightly higher share of horizontal equity concerns in Treatment Low-Corr likely helps explain the lower support for tagging observed in this treatment.

Comparing the Immutable and Mutable treatments reveals a higher proportion of horizontal equity concerns in the Immutable treatment (65% in Treatment Immutable and 47% in Treatment Mutable), while participants in the Mutable treatment more often emphasize vertical equity considerations (32% in Treatment Immutable and 44% in Treatment Mutable). However, these differences do not account for the observed null effect in support across these two treatments, suggesting the presence of other concerns that increase support for tagging in Treatment Immutable.

Finally, comparing Corr-Needs and Uncorr-Needs treatments shows that participants in Corr-Needs more frequently mention supporting needy individuals (25% in Treatment Uncorr-Needs and 34% in Treatment Corr-Needs), whereas those in Uncorr-Needs more often cite horizontal equity concerns (61% in Treatment Uncorr-Needs and 52% in Treatment Corr-

Needs). These patterns are consistent with the higher support observed in the Corr-Needs treatment.

Summary: Overall, participants express horizontal equity concerns, as hypothesized in the literature; however, vertical equity concerns are also prominent. In particular, some participants view the groups as equals and emphasize, consistent with horizontal equity, that individuals with the same income should face the same tax burden regardless of group affiliation. Others view the groups as meaningfully different in terms of income, needs, or disadvantages, and argue, consistent with vertical equity, for supporting groups with lower incomes, higher needs, or disadvantages. These perceptions offer a possible explanation for the variation in support across individuals and across treatments.

4.4 Underlying Concerns: Fairness Concerns

The open-ended responses documented in the previous section highlight fairness—particularly horizontal and vertical equity—as the most salient first-order concern. In this section, we formalize the role of fairness concerns and test whether they account for heterogeneity in support across individuals and treatments in the vignette experiment.

We conceptualize fairness through welfare weights (see Section 2) and measure them using an incentivized bonus allocation task adapted from Capozza & Srinivasan (2024). After the vignette experiment, participants chose whether to allocate a $500 bonus to an individual from the tagged group, to one from the untagged group, or to either (indicating indifference). Allocating the bonus to the tagged group reflects progressive welfare weights (since the tagged group has a lower average income), allocating it to the untagged group reflects regressive weights, and indifference implies equal weights across groups.[17]

Pooling across all treatments, 48% of participants have progressive weights, 42% have equal weights, and 10% have regressive weights (Appendix Figure A14). We also find variation in welfare weights across treatments. We exclude participants with regressive weights from the main analysis, since prior work suggests that some regressive choices are driven by mistakes (Capozza & Srinivasan 2024); however, our results do not hinge on this exclusion.

Variation Across Individuals: Conditional on having progressive welfare weights, the share of participants supporting tagging is 60%. In contrast, among those with equal weights, only 16% support tagging.[18] Figure 4 presents this association across treatments. We find a

[17]The welfare weight assigned to an individual from a group is assumed to represent the average weight assigned to that group.

[18]Among the 10% of participants with regressive weights, 22% support tagging.

statistically significant association between support for tagging and welfare weights in each of the eight treatments (chi-square tests; $p < 0.05$). To summarize, the variation in welfare weights across individuals predicts the variation in support for tagging across individuals.

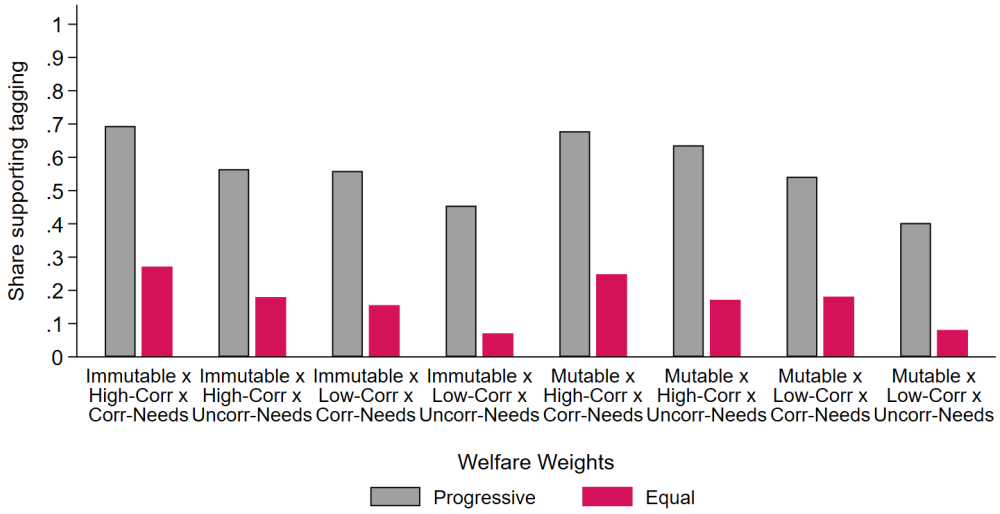


Figure 4: Share Supporting Tagging by Welfare Weights and Treatment

Notes: The figure presents the share of participants supporting tagging in the vignette experiment, by welfare weights and treatment. A participant has progressive welfare weights if they allocate the bonus to the individual from the tagged group, and equal weights if they are indifferent between allocating the bonus to the individual from the tagged group and the untagged group. Participants with regressive weights are excluded.

Variation Across Treatments: Figure 5, which presents the share of participants with progressive weights by treatment, closely parallels the pattern in support for tagging observed in Figure 3. Participants assign higher welfare weights in Treatment High-Corr relative to Treatment Low-Corr, and in Treatment Corr-Needs relative to Uncorr-Needs—a pattern that aligns with the observed treatment effects on support for tagging.[19] Both treatment differences are statistically significant (Appendix Table A8). Participants' welfare weights are very similar across the Mutable and Immutable treatments, and the difference is not statistically significant—mirroring the null effect on support for tagging observed across these treatments (Appendix Table A8). To summarize, the variation in welfare weights across treatments predicts the variation in support for tagging across treatments.

[19] A mediation analysis shows that welfare weights explain part of the treatment variation in support: they account for 45% of the difference in support between Treatments High-Corr and Low-Corr, and 33% of the difference between Treatments Corr-Needs and Uncorr-Needs (Appendix Table A9).

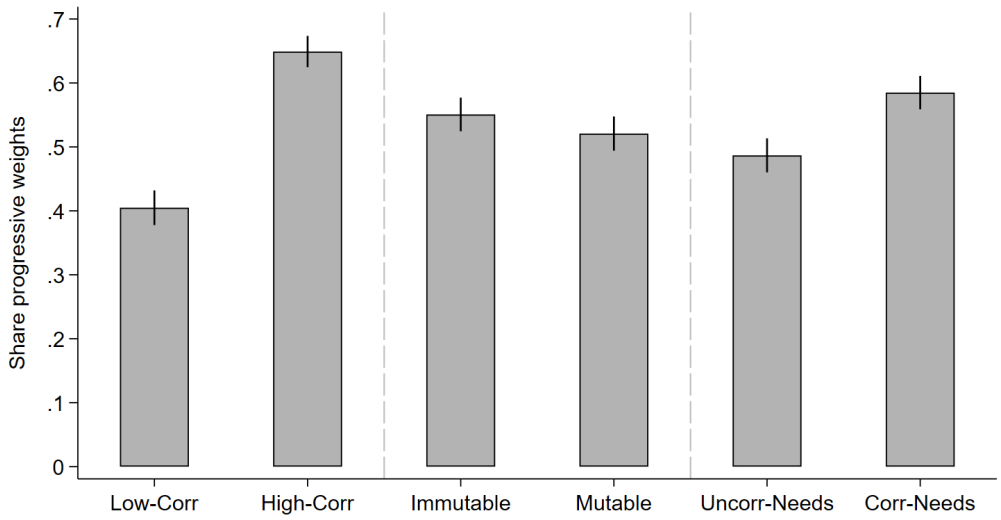


Figure 5: Share Assigning Progressive Welfare Weights by Treatment

Notes: The figure presents the share of participants assigning progressive welfare weights by treatment. A participant has progressive welfare weights if she allocates the bonus to the individual from the tagged group, and equal weights if she is indifferent between allocating the bonus to the individual from the tagged group and the untagged group. Participants with regressive weights are excluded.

Summary: Overall, the results in this section suggest that fairness concerns (capturing horizontal and vertical concerns)—measured by welfare weights—predict support for tagging, helping to explain variation across individuals and treatments. Yet welfare weights do not fully account for the observed variation, suggesting that other concerns may also be driving support. For example, participants with progressive weights (implying a preference to increase between-group redistribution) may oppose tagging because of concerns about efficiency, implementation, scope of government, or privacy. Conversely, participants with equal weights (implying a preference to maintain the status quo incomes) may support tagging for efficiency reasons, such as to reduce labor distortions or increase government revenue.

4.5 Underlying Concerns: Overarching Concerns

Having identified the role of fairness concerns, we study other overarching concerns, including efficiency, scope of government, implementation, and privacy.

Variation Across Individuals: Figure 6 reports coefficients from linear regressions where the dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise, and the explanatory variables are binary indicators capturing various mechanisms and background characteristics. Variable definitions are provided in Appendix Sec-

tion B. Appendix Table A10 presents the corresponding regression tables. Fairness concerns are the most important predictor of support for tagging: participants who view incomes from tagging as fair are 28 percentage points (pp) more likely to support tagging than those who are neutral or view them as unfair.[20] Efficiency concerns also matter, but less so. Support is 5 pp higher among those who believe tagging increases government revenue, with a similar effect for those who think tagging raises high-income labor supply. Surprisingly, those who believe tagging increases evasion are about 6 pp more likely to support tagging. Participants who perceive the government as inefficient are 3.5 pp less likely to support tagging. In addition to fairness and efficiency, other concerns also matter: support for tagging is 7 pp higher among those who think government should do more (scope of government), 6 pp higher among those with higher trust in government (implementation concerns), and 5 pp lower among those concerned about privacy (privacy concerns). Demographics explain additional variation. Support is 4 pp higher among above-median educated individuals, while it is 6-9 pp lower among above-median aged individuals, above-median income individuals, and Republicans. We find no significant gender effects.

Identifying the relative importance of concerns is challenging because the variables within each concern and across concerns may be correlated. To address this issue, we employ a Shapley value decomposition that identifies the contribution of each concern to explaining the overall variation (R^2) (Lipovetsky 2021).[21] The R^2 of our regression is 0.13. The decomposition shows that fairness concerns account for 62% of the explained variation, efficiency for 18%, scope of government for 13%, privacy for 5%, and implementation concerns for 2%. The results are similar using a continuous measure for our outcome variable and explanatory variables (Appendix Table A11).

Variation Across Treatments: For the two treatment dimensions where we detect a treatment effect, we examine which concerns account for the treatment variation using a mediation analysis. In particular, we conduct a Gelbach decomposition (Gelbach 2016), which identifies the share of the treatment difference attributable to the various concerns,

[20] We use a survey measure of fairness rather then the incentivized measure presented in the previous section since the former is more comparable to the survey measures capturing other concerns. Capozza & Srinivasan (2024) shows that variation in quantitative measures of welfare weights is accounted for by this survey measure of fairness, suggesting that survey measure likely captures a broad set of underlying normative ideals.

[21] The Shapley value, derived from cooperative game theory, provides a principled way to assess the marginal contribution of each concern to the explained variance (R^2) in our model. Intuitively, this approach considers all possible combinations of explanatory variables and calculates the incremental contribution of each when added to different subsets of variables. By averaging these contributions across all permutations, the Shapley value fairly allocates the R^2 among the concerns, accounting for their interdependence.

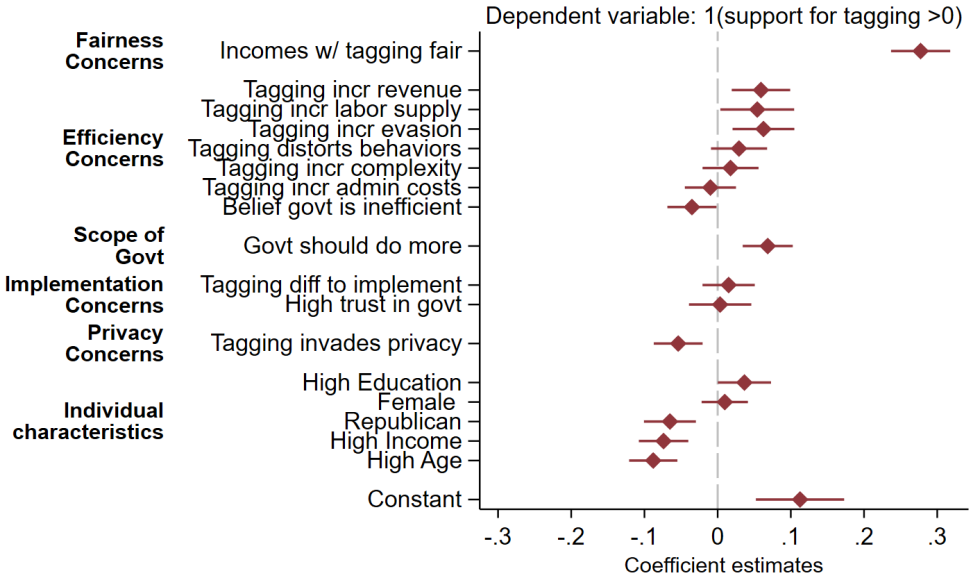


Figure 6: Support for Tagging and Mechanisms

Notes: The figure presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. The explanatory variables include mechanisms, background characteristics, and treatment indicators (omitted). All explanatory variables capturing mechanisms are binary variables (see Appendix Section B for the definitions). *Republican* is an indicator for Republicans. *High Age* equals 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. Bars represent the 95% confidence intervals based on HC3 standard errors.

while controlling for potential correlations among the concerns.[22]

[23] Table 4 presents the results. Fairness concerns explain the largest share of the treatment differences (3% of the High-Corr vs. Low-Corr difference and 9% of the Corr-Needs vs. Uncorr vs. Corr-Needs difference), followed by efficiency concerns (1% and 2%, respectively). Other concerns, including scope of government, privacy, and implementation, contribute negligibly.

[22]The Gelbach decomposition compares the treatment coefficient from a regression without mechanism variables to that from a regression including them. For example, the High-Corr coefficient declines from 0.227 (col. 1) to 0.214 (col. 2), a reduction of 0.013 or about 5.7% of the initial effect, implying that mechanisms collectively explain roughly 6% of the treatment difference (Appendix Table A12). Similarly, in the Corr-Needs specification, the treatment effect declines from 0.124 (col. 3) to 0.115 (col. 4), a reduction of 0.009 or about 7.3%, indicating that mechanisms account for approximately 7% of that treatment difference. The decomposition then allocates these explained shares across the mechanism groups.

[23]The Gelbach decomposition involves comparing the treatment coefficient from a regression without mechanism variables to that from a regression including them. The Low-Corr vs. High-Corr treatment difference declines from 0.227 to 0.214, implying that mechanisms explain about 6% of the treatment difference (Appendix Table A12). The decomposition then allocates this share across the mechanisms. Similarly, the Uncorr-Needs vs. Corr-Needs difference declines from 0.124 to 0.115, implying that the mechanisms account for about 7% of the treatment difference (Appendix Table A12).

Table 4: Decomposition of Treatment Effects

	(1)	(2)
Fairness Concerns	0.03	0.09
Efficiency concerns	0.01	0.01
Scope of govt	0.00	0.01
Privacy concerns	0.00	0.00
Implementation concerns	0.00	0.00
Share unexplained	0.94	0.93
Treatment	High-Corr	Corr-Needs

Notes: The table presents a covariate decomposition of treatment effects on support for tagging in the vignette experiment, following Gelbach (2016). Each entry shows the fraction of the treatment effect attributable to the indicated factor. The last row reports the unexplained share. Rows sum to one, up to rounding.

For the null effect between the Mutable and Immutable treatments, we examine whether this result is moderated by various mechanisms. Specifically, we estimate heterogeneous treatment effects to assess whether the treatment effect differs among participants with varying beliefs about, and preferences for, tagging. Across all specifications, we do not find statistically significant heterogeneous effects with respect to any of the mechanisms (Appendix Figure A15). The results related to efficiency are particularly noteworthy: they indicate that the difference between the Mutable and Immutable treatments remains small even among participants who recognize that tagging can distort choices or increase evasion—two margins along which, in theory, immutable tags should be preferred. These results provide suggestive evidence that we would likely observe only limited differences between the two treatments, even in a setting where participants fully understood and internalized the efficiency gains from immutable tags.

Summary: To summarize, fairness concerns are the most important predictor of support for tagging, explaining the largest share of variation in support for tagging both across individuals and treatments. Efficiency concerns also matter, but less so. Other concerns, such as scope of government, implementation, and privacy, help account for individual-level variation but contribute little to differences across treatments. Demographic characteristics also explain a part of the individual variation: higher-educated individuals express support for tagging, while older individuals, higher-income individuals, and Republicans express lower support.

5 Theory, Practice, and Citizens' Support

In this concluding section, we bring together insights from theory, existing policies, and our experimental findings. We ask which tags proposed in the literature are likely to be socially acceptable and whether the tags used in practice reflect citizens' views. To identify the tags discussed in the literature, we search for studies that have explored the efficiency benefits of employing specific tags using theory or simulations. Table 5 presents the tags documented in the literature, which broadly relate to individuals' family structure, personal attributes, education and occupation, health status, and economic status. To identify the tags used in practice, we construct a novel dataset of the tags used in the U.S. tax system.[24] Column (1) of Table 5 lists various groups and Column (2) indicates the relative tax burden of these groups compared to the rest of the population—it can be higher, lower, the same, or dependent on other factors. Appendix Table A1 provides detailed explanations of the tax burden of the groups. We find that the U.S. tax system employs a wide range of tags.

Theory and Practice: There are three insights from exploring the tax discussed in the literature. First, several tags highlighted in the literature are also implemented in the U.S. tax system. These include the presence of children, marital status, race and ethnicity, age, place of residence, sector of employment, disability status, and visual impairment status. However, the extent of tagging for tags on age and sector of employment—in terms of the number of groups—is limited, relative to the benchmarks considered in the literature (e.g., Weinzierl 2011, Gaubert et al. 2021). Moreover, some of the groups that receive a tax break in practice, such as farmers and fishermen or those with children, earn higher incomes, highlighting other policy goals, such as incentivizing socially desirable behaviors or correcting externalities. Some of these goals have been discussed in literature (e.g., Lockwood et al. 2017, Farhi & Werning 2013). Second, several tags proposed in the literature—such as those based on height, genes, gender, level of education, and accidental bequests—are not currently implemented by the U.S. tax system. In our experiment, we show that participants have low support for taxation based on gender and education. Future research can explore whether the other tags would be socially acceptable. In light of our results, these tags would be socially acceptable if they are a very good signal of ability or related to needs. Finally, although most of the tags implemented by the U.S. tax system have been discussed in the literature, some remain unexplored. These include dependents, student status, homeownership status, self-

[24] We identified the tax burden of various groups using the documentation of the Internal Revenue Service (IRS), the federal tax authority. We then hired a freelancer specializing in tax preparation to verify our findings. Individuals' tax burden is determined by tax rates, deductions, exemptions, tax credits, and other tax rules. We exclude in-kind transfers, such as tax filing assistance.

employment status, and veteran status.[25] To summarize, there's scope for expanding tagging as well as scope for future research to identify tags that are both economically beneficial and socially acceptable.

Practice and Citizens' Support: Do the tax policies in place reflect citizens' views? To answer this question, we compare participants' support for providing a tax break to various groups with the groups that receive a tax break in practice. The groups that receive strong public support—those with dependents, disabilities, visual impairments, and living in low-income regions—also receive a tax break in practice, while groups that receive limited public support generally do not receive a tax break in practice.[26] Some groups that were not included in our survey—such as adopted children, widows, Native Americans, seniors, minors, clergy, farmers and fishermen, the self-employed, students, veterans, and homeowners—do receive tax breaks in practice. Future work could examine citizens' support for these groups. We hypothesize low support for some of these, such as Native Americans, clergy, and students, given our findings that people do not support race-based, sector-based, and education-based taxation. To summarize, citizens' preferences and existing tax policies are broadly aligned, with a few exceptions.

6 Discussion

Tagging—conditioning taxes jointly on income and characteristics—can increase efficiency but is widely believed to lack public support because it conflicts with the principle of horizontal equity, which stipulates treating equals equally. We test this claim using an online vignette experiment with a U.S general population sample and report three sets of results that challenge the view.

First, support for tagging is heterogeneous across both individuals and tags. The median respondent supports about one-third of tags, but there is wide dispersion across individuals. Likewise, although support for many tags is low, several tags receive substantial support. This pattern indicates that opposition to tagging is not universal; attitudes appear to be shaped by both the characteristics of tags and individual beliefs and preferences. Focusing on the variation across tags using a vignette experiment, we find higher support when the

[25] Homeownership has been studied in relation to tags correlated with individuals' labor supply elasticities (Karabarbounis 2016). Student status has been widely studied in the context of human capital accumulation (Bovenberg & Jacobs 2005).

[26] One exception is marital status: despite limited support for tax breaks for the unmarried, the tax code conditions on marital status. However, in practice, the tax burden of unmarried individuals can be higher or lower, depending on the difference in incomes of the two earners.

tag is a strong signal of ability or relates to consumption needs, consistent with theoretical prescriptions. However, the immutability of the tag does not predict support, contrasting the theoretical prescription of using immutable tags.

Second, vertical equity also matters alongside horizontal equity. Evidence from open-ended responses and from a direct measurement of horizontal and vertical equity via welfare weights shows that both concerns help account for the variation in support across individuals and tags. Conceptually, if the tagged and untagged groups have the same average welfare weights, horizontal equity implies no tagging; if their average weights differ, vertical equity supports redistribution from the group with the lower to the higher welfare weights, thereby rationalizing tagging.

Third, while fairness concerns are important, other considerations also shape support. Efficiency concerns also predict support, though less so. Other concerns, such as the scope of government, implementation, and privacy, account for individual-level variation but explain little of the differences across treatments.

While the canonical utilitarian model has two goals—improving efficiency and between-group redistribution—our results suggest that participants are likely guided primarily by redistributive motives. Given the importance of such fairness concerns, we incorporate them into our theoretical model to characterize when tagging is socially acceptable. We model social acceptability via the welfare weights assigned to the tagged and untagged groups. The optimal tax differentiation between groups is increasing in the difference between their average welfare weights.

Finally, we synthesize our experimental results, theory, and practice. There is scope for expanding tagging based on existing research, as well as scope for future research to identify tags that are both economically beneficial and socially acceptable. Furthermore, citizens' preferences and existing tax policies are broadly aligned, with a few exceptions.

Future Research: There are several directions for future research. First, future studies can examine support for tagging when the tag is correlated with individuals' preferences or their elasticity of taxable income (ETI), which can lead to efficiency benefits (e.g., Alesina et al. 2011, Karabarbounis 2016). Our study included some tags that are correlated with labor supply elasticity, such as gender and the presence of dependents. Prior evidence from controlled experiments (Sausgruber & Tyran 2014) and vignette studies (e.g., Saez & Stantcheva 2016) suggests that individuals are reluctant to base taxes on preferences. We hypothesize that support may be higher for tags correlated with preferences or ETI if these tags are also associated with individuals' needs.

Second, future research can study Pareto improving tagging, which involves differential

taxation across groups, but lower tax rates for all individuals relative to the status quo. While Mankiw & Weinzierl (2010) find modest gains from Pareto-improving reforms for height-based taxation, future work could examine the gains for other tags. Furthermore, future research can explore citizens' support for Pareto improving tagging. However, because Pareto improving tagging does not resolve horizontal equity concerns, support for tagging may remain limited.

Third, future research could examine tagging with different policy instruments, such as tax credits, deductions, or commodity taxes. Under horizontal equity, which requires treating equals equally, varying the instrument effectively varies how citizens interpret "treating equally," holding fixed their perceptions about the equality of groups. Future work could also analyze tagging implemented through a set of policies rather than a single tax instrument, consistent with existing U.S. tax and transfer policies revealing welfare weights that depend on the age of beneficiaries (Hendren & Sprung-Keyser 2020).

Table 5: List of Tags

Tag and Groups	Tax Burden	Research papers
	Family Structure	
Children		Cigno (2001), Balestrino et al. (2002), Blumkin et al. (2015)
- Presence of children	Lower	
- Presence of adopted children	Lower	
Dependents		
- Presence of dependents*	Lower	
Marital status		Liebman & Ramsey (2019), Bastani (2013)
- Married*	Depends	
- Widow(er)	Lower	
	Personal Attributes	
Race and ethnicity		Cushing & McGarvey (2003), Blumkin et al. (2009), Weinzierl (2012)
- Native Americans	Lower	
- Ethnic/Racial minorities*	Same	
Height		Mankiw & Weinzierl (2010)
- Short individuals	Same	

Continued on next page

Table 5: List of Tags (continued)

Tag	Tax Burden	Research Papers
Genes - Genes correlated with earnings	Same	Logue & Slemrod (2008)
Gender - Women*	Same	Cushing & McGarvey (2003), Bastani (2013), Cremer et al. (2010), Weinzierl (2012)
Age - Minors - Young* - Seniors	Higher Same Lower	Kremer (2002), Blomquist & Micheletto (2008), Bastani et al. (2013), Weinzierl (2011), Kanbur & Tuomala (2016), Heathcote et al. (2020)
Place of residence - Low-income regions*	Lower	Obara (2019), Kessing et al. (2020), Gaubert et al. (2021)
Education and Occupation		
Level of education - Low-educated*	Same	Blumkin & Sadka (2008)
Sector of employment - Clergy - Farmers and fisherman - Not service sector*	Lower Lower Same	Gomes et al. (2018)
Self-employment status - Self-employed	Higher	
Student status - Students	Lower	
Veteran status - Veterans	Lower	
Health Status		
Disability status - Disabled*	Lower	Jacquet (2014)
Visual impairment status - Visually impaired*	Lower	Weinzierl (2012)
Economic Status		

Continued on next page

Table 5: List of Tags (continued)

Tag	Tax Burden	Research Papers
Homeownership status		
- Homeowners	Lower	
Accidental bequests		Cremer et al. (2012)
- Bequest beneficiaries	Same	

Column (1) presents various tags and the groups that can be defined based on these tags. Column (2) indicates the relative tax burden of these groups compared to the rest of the population—it can be higher, lower, the same, or dependent on other factors. Column (3) indicates papers that focus on the tags. Groups marked with stars are included in the experiment.

References

Akerlof, G. A. (1978), 'The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning', *The American economic review* **68**(1), 8–19.

Alesina, A., Ichino, A., Karabarbounis, L. & Loungani, P. (2011), 'Gender-based taxation and the division of family chores', *American Economic Review* **101**(3), 414–418.

Andersson, O., Holm, H. J., Tyran, J.-R. & Wengström, E. (2016), 'Deciding for others reduces loss aversion', *Management Science* **62**(1), 29–36.

Balestrino, A., Cigno, A. & Pettini, A. (2002), 'Endogenous fertility and the design of family taxation', *International Tax and Public Finance* **9**, 175–193.

Bastani, S. (2013), 'Gender-based and couple-based taxation', *International tax and public finance* **20**, 653–686.

Bastani, S., Blomquist, S. & Micheletto, L. (2013), 'The welfare gains of age-related optimal income taxation', *International Economic Review* **54**(4), 1219–1249.

Berg, K. (2024), 'Why taxes need not treat equals equally', *Erasmus Journal for Philosophy and Economics* **17**(2), 1–22.

Berg, K. (2025), 'Revealing inequality aversion from tax policy and the role of non-discrimination', *Economica* **92**(366), 483–506.

Bergstrom, K. & Dodds, W. (2021), 'Using labor supply elasticities to learn about income inequality: the role of productivities versus preferences', *American Economic Journal: Economic Policy* **13**(3), 28–62.

Blesse, S., Buhlmann, F. & Dörrenberg, P. (2019), 'Do people really want a simple tax system? Evidence on preferences towards income tax simplification'. *ZEW Discussion Paper no. 19-058*.

Blomquist, S. & Micheletto, L. (2008), 'Age-related optimal income taxation', *Scandinavian Journal of Economics* **110**(1), 45–71.

Blumkin, T., Margalioth, Y. & Sadka, E. (2009), 'Incorporating affirmative action into the welfare state', *Journal of Public Economics* **93**(9-10), 1027–1035.

Blumkin, T., Margalioth, Y. & Sadka, E. (2015), 'The re-distributive role of child benefits revisited', *International Tax and Public Finance* **22**, 476–501.

Blumkin, T. & Sadka, E. (2008), 'A case for taxing education', *International Tax and Public Finance* **15**, 145–163.

Boadway, R. & Pestieau, P. (2006), 'Tagging and redistributive taxation', *Annales d'Economie et de Statistique* pp. 123–147.

Bovenberg, A. L. & Jacobs, B. (2005), 'Redistribution and education subsidies are siamese twins', *Journal of Public Economics* **89**(11-12).

Capozza, F. & Srinivasan, K. (2024), 'Who should get money? Estimating welfare weights in the U.S.'. *CESifo Working Paper no. 11086*.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø. & Tungodden, B. (2007), 'The pluralism of fairness ideals: An experimental approach', *American Economic Review* **97**(3), 818–827.

Cappelen, A. W., Moene, K. O., Sørensen, E. Ø. & Tungodden, B. (2013), 'Needs versus entitlements—an international fairness experiment', *Journal of the European Economic Association* **11**(3), 574–598.

Cigno, A. (2001), 'Comparative advantage, observability, and the optimal tax treatment of families with children', *International Tax and Public Finance* **8**, 455–470.

Cremer, H., Dellis, A. & Pestieau, P. (2003), 'Family size and optimal income taxation', *Journal of Population Economics* **16**, 37–54.

Cremer, H., Gahvari, F. & Lozachmeur, J.-M. (2010), 'Tagging and income taxation: theory and an application', *American Economic Journal: Economic Policy* **2**(1), 31–50.

Cremer, H., Gahvari, F. & Pestieau, P. (2012), 'Accidental bequests: A curse for the rich and a boon for the poor', *The Scandinavian Journal of Economics* **114**(4), 1437–1459.

Cushing, M. J. & McGarvey, M. G. (2003), 'The efficiency of race-and gender-targeted income transfers', *Public Finance Review* **31**(5), 455–486.

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

Elias, J. J., Lacetera, N. & Macis, M. (2019), 'Paying for kidneys? a randomized survey and choice experiment', *American Economic Review* **109**(8), 2855–2888.

Farhi, E. & Werning, I. (2013), 'Insurance and taxation over the life cycle', *Review of Economic Studies* **80**(2), 596–635.

Fehr, E., Epper, T. & Senn, J. (2024), 'Social preferences and redistributive politics', *Review of Economics and Statistics* pp. 1–45.

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

Fleurbaey, M. & Maniquet, F. (2018), 'Optimal income taxation theory and principles of fairness', *Journal of Economic Literature* **56**(3), 1029–1079.

Galbiati, R. & Vertova, P. (2006), 'Horizontal equity', *Economica* **73**(290), 771–791.

Gaubert, C., Kline, P. M. & Yagan, D. (2021), Place-based redistribution. National Bureau of Economic Research.

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

Gomes, R., Lozachmeur, J.-M. & Pavan, A. (2018), 'Differential taxation and occupational choice', *The Review of Economic Studies* **85**(1), 511–557.

Haaland, I. K., Roth, C., Stantcheva, S. & Wohlfart, J. (2025), 'Understanding economic behavior using open-ended survey data', *Journal of Economic Literature (forthcoming)* . Forthcoming.

Heathcote, J., Storesletten, K. & Violante, G. L. (2020), 'Optimal progressivity with age-dependent taxation', *Journal of Public Economics* **189**, 104074.

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

Immonen, R., Kanbur, R., Keen, M. & Tuomala, M. (1998), 'Tagging and taxing: The optimal use of categorical and income information in designing tax/transfer schemes', *Economica* **65**(258), 179–192.

Jacquet, L. (2014), 'Tagging and redistributive taxation with imperfect disability monitoring', *Social Choice and Welfare* **42**, 403–435.

Jäger, S., Roth, C., Roussille, N. & Schoefer, B. (2024), 'Worker beliefs about outside options', *The Quarterly Journal of Economics* **139**(3), 1505–1556.

Kanbur, R. & Tuomala, M. (2016), 'Groupings and the gains from tagging', *Research in Economics* **70**(1), 53–63.

Kaplow, L. (1989), 'Horizontal equity: measures in search of a principle', *National Tax Journal* **42**(2), 139–154.

Karabarbounis, M. (2016), 'A road map for efficiently taxing heterogeneous agents', *American Economic Journal: Macroeconomics* **8**(2), 182–214.

Kessing, S. G., Lipatov, V. & Zoubek, J. M. (2020), 'Optimal taxation under regional inequality', *European Economic Review* **126**, 103439.

Konow, J. (2001), 'Fair and square: the four sides of distributive justice', *Journal of Economic Behavior & Organization* **46**(2), 137–164.

Kremer, M. (2002), 'Should taxes be independent of age? *Working Paper*.

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.

Liebman, J. & Ramsey, D. (2019), 'Independent taxation, horizontal equity, and return-free filing', *Tax Policy and the Economy* **33**(1), 109–130.

Lipovetsky, S. (2021), 'Game theory in regression modeling: A brief review on shapley value regression', *Model Assisted Statistics and Applications* **16**(2), 165–168.

Liscow, Z. & Pershing, A. (2022), 'Why is so much redistribution in-kind and not in cash? Evidence from a survey experiment', *National Tax Journal* **75**(2), 313–354.

Lockwood, B. B., Nathanson, C. G. & Weyl, E. G. (2017), 'Taxation and the allocation of talent', *Journal of Political Economy* **125**(5), 1635–1682.

Lockwood, B. B. & Weinzierl, M. (2015), 'De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution', *Journal of Public Economics* **124**, 74–80.

Logue, K. D. & Slemrod, J. (2008), 'Genes as tags: The tax implications of widely available genetic information', *National Tax Journal* **61**(4), 633–659.

Mankiw, N. G. & Weinzierl, M. (2010), 'The optimal taxation of height: A case study of utilitarian income redistribution', *American Economic Journal: Economic Policy* **2**(1), 155–176.

Mankiw, N. G., Weinzierl, M. & Yagan, D. (2009), 'Optimal taxation in theory and practice', *Journal of Economic Perspectives* **23**(4), 147–174.

Milligan, K. (2005), 'Subsidizing the stork: New evidence on tax incentives and fertility', *Review of Economics and statistics* **87**(3), 539–555.

Mirrlees, J. A. (1971), 'An exploration in the theory of optimum income taxation', *The review of economic studies* **38**(2), 175–208.

Obara, T. (2019), 'Differential income taxation and tiebout sorting', *FinanzArchiv: Public Finance Analysis* **75**(1), 1–38.

Parsons, D. O. (1996), 'Imperfect 'tagging'in social insurance programs', *Journal of Public Economics* **62**(1-2), 183–207.

Pestieau, P. & Racionero, M. (2015), 'Tagging with leisure needs', *Social Choice and Welfare* **45**, 687–706.

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

Saez, E. & Stantcheva, S. (2016), 'Generalized social marginal welfare weights for optimal tax theory', *American Economic Review* **106**(01), 24–45.

Salanié, B. (2002), 'Optimal demogrants with imperfect tagging', *Economics Letters* **75**(3), 319–324.

Sausgruber, R. & Tyran, J.-R. (2014), 'Discriminatory taxes are unpopular—even when they are efficient and distributionally fair', *Journal of Economic Behavior & Organization* **108**, 463–476.

Slemrod, J. (2025), 'Tax privacy', *Journal of Economic Perspectives* **39**(1), 205–224.

Stantcheva, S. (2021), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 2309–2369.

Viard, A. D. (2001*a*), 'Optimal categorical transfer payments: The welfare economics of limited lump-sum redistribution', *Journal of Public Economic Theory* **3**(4), 483–500.

Viard, A. D. (2001*b*), 'Some results on the comparative statics of optimal categorical transfer payments', *Public Finance Review* **29**(2), 148–180.

Weinzierl, M. (2011), 'The surprising power of age-dependent taxes', *The Review of Economic Studies* **78**(4), 1490–1518.

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. C. (2012), 'Why do we redistribute so much but tag so little? The principle of equal sacrifice and optimal taxation'. *NBER Working Paper no. 18045*.

ONLINE APPENDIX

People's Preferences for Tagging

Julien Senn
Krishna Srinivasan

Table of Contents

A Proofs

Proof of Proposition 1

Setup: This section provides proofs of the propositions in Section 2. We begin by describing the setup, which closely follows Mankiw & Weinzierl (2010). The social planner's objective is to set optimal taxes to maximize a utilitarian social welfare function that is uniform and linear in individuals' utilities. As shown by Mankiw & Weinzierl (2010) and others, the planner's problem can be decomposed into two separable problems: (i) setting optimal taxes within groups and (ii) setting optimal aggregate transfers between groups. The planner's problem is as follows

$$\max_{c,y,r;} \sum_{g\in t,u} \frac{1}{2} \sum_{i=1}^{I} \pi_{g,i} u\left(c_{g,i}, \frac{y_{g,i}}{w_i}\right), \tag{A1}$$

subject to two within-group feasibility constraints

$$\sum_{i=1}^{I} \pi_{g,i} \left(y_{g,i} - c_{g,i}\right) \geq r_g, \tag{A2}$$

a between-group budget constraint

$$r_t + r_u = 0, \tag{A3}$$

and incentive constraints for each individual

$$u\left(c_{g,i}, \frac{y_{g,i}}{w_i}\right) \geq u\left(c_{g,j}, \frac{y_{g,j}}{w_i}\right), \tag{A4}$$

for all $j \neq i \in \{1, ..., I\}$ for each individual of group g with wage w_i.

Proposition 1 states: *When an observable and immutable personal characteristic is correlated with earning ability, it is optimal to condition taxes on that characteristic.*

Proof: Let λ_g denote the Lagrange multiplier on the within-group feasibility constraints (A2), and μ the multiplier on the between-group budget constraint (A3). The Lagrangian is

$$\mathcal{L} = \frac{1}{2} \sum_{g} \sum_{i=1}^{I} \pi_{g,i}\, u\left(c_{g,i}, \frac{y_{g,i}}{w_i}\right) + \sum_{g} \lambda_g \left[\sum_{i=1}^{I} \pi_{g,i}(y_{g,i} - c_{g,i}) - r_g\right] \tag{A5}$$

$$+\mu(r_t + r_u) + (\text{IC multipliers}). \tag{A6}$$

The *IC terms* (omitted for brevity) in the Lagrangian collect the multipliers and expressions attached to the incentive compatibility constraints (A4). Differentiating $\mathcal{L}$ with respect to r_g gives

$$\frac{\partial \mathcal{L}}{\partial r_g} = -\lambda_g + \mu = 0 \quad \Rightarrow \quad \lambda_t = \lambda_u = \mu, \tag{A7}$$

implying equality of the shadow value of resources across groups. Differentiating $\mathcal{L}$ with respect to $c_{g,i}$ gives

$$\frac{\partial \mathcal{L}}{\partial c_{g,i}} = \frac{1}{2}\pi_{g,i}\, u_c(c_{g,i}, \ell_{g,i}) - \lambda_g \pi_{g,i} + (\text{IC multiplies}) = 0. \tag{A8}$$

Here, the IC terms collect the derivatives of the incentive-compatibility constraints. When we sum (A8) over i within a group, these IC terms cancel pairwise due to the symmetric structure of the IC constraints. The resulting condition for group g is

$$\sum_{i=1}^{I} \left[\frac{1}{2}\pi_{g,i}\, u_c(c_{g,i}, \ell_{g,i}) - \lambda_g \pi_{g,i} \right] = 0 \tag{A9}$$

$$\Rightarrow \quad \lambda_g = \frac{1}{2} \sum_{i=1}^{I} \pi_{g,i}\, u_c(c_{g,i}, \ell_{g,i}). \tag{A10}$$

Combining Equations (A7) and (A10) yields

$$\sum_{i=1}^{I} \pi_{t,i}\, u_c(c_{t,i}, \ell_{t,i}) = \sum_{i=1}^{I} \pi_{u,i}\, u_c(c_{u,i}, \ell_{u,i}). \tag{A11}$$

This completes the proof of Proposition 1.

Proof of Proposition 2

Setup:

Definition 1 (Strength of tag-ability correlation) *Let $\alpha \in [0,1]$ parameterize the* strength of the correlation *between the tag and ability w. An increase in α means that*

$$\frac{\partial \mathbb{E}_t[w]}{\partial \alpha} < 0 \quad \textit{and} \quad \frac{\partial \mathbb{E}_u[w]}{\partial \alpha} > 0.$$

We assume allocations satisfy the Spence–Mirrlees single-crossing condition, which states

that the marginal rate of substitution between consumption and before-tax income is decreasing in w. Together with incentive compatibility constraints (A4), this implies the standard monotonicity of optimal allocations

Assumption 1 (Monotonicity) *Under the Spence–Mirrlees single-crossing condition and incentive compatibility* (A4)*, the optimal allocation satisfies*

$$y_g(w_i) \text{ and } c_g(w_i) \text{ are (weakly) increasing in } w_i, \qquad g \in \{t, u\}.$$

Proposition 2 states: *The optimal between-group transfer is increasing in the strength of correlation of the tag with ability.* When α increases, $\partial r/\partial \alpha > 0$.

Proof: Starting from an optimal allocation and between-group transfer that satisfies Equation (A11), suppose we increase α as in Definition 1. Using assumption (1), the tagged group's average consumption falls while the untagged group's average consumption rises. Since u_c is strictly decreasing in consumption, the tagged group's average marginal utility of consumption increases and the untagged group's decreases, violating Equation (A11). Consequently, the planner must increase the between-group transfer to the tagged group to restore equality of average marginal utility of consumption and satisfy Equation (A11). Hence, $\partial r/\partial \alpha > 0$.

Proof of Proposition 3

Setup:

Definition 2 (Mutability) *Let $g \in \{t, u\}$ denote true group membership and $\hat{g} \in \{t, u\}$ the reported group. The tag is* immutable *if reporting is constrained to be truthful for all individuals ($\hat{g}_i = g_i$ for all i). The tag is one-sided mutable if individuals in group u may report $\hat{g}_i = t$ even if $g_i = u$, while individuals in group t must report truthfully: $\hat{g}_i = t$ whenever $g_i = t$.*

Proposition 3 states: *The optimal between-group transfer is higher when the tag is immutable than when it is mutable.*

Proof: For immutable tags, the planner's problem is as described above. With one-sided mutability (Definition 2), individuals in the untagged group may misreport their group status to receive the transfer given to the tagged group. To prevent such misreporting, we require

an additional incentive compatibility constraint ensuring that no untagged individual prefers the tagged allocation:

$$u\left(c_{u,i}, \frac{y_{u,i}}{w_i}\right) \geq u\left(c_{t,i}, \frac{y_{t,i}}{w_i}\right) \qquad \text{for all } i \text{ with } g_i = u. \tag{A12}$$

We do not require a symmetric constraint for the tagged group $g = t$ because they report truthfully. Since one-sided mutability introduces this additional constraint, the feasible set of allocations under mutability is a weak subset of the feasible set under immutability. Let r^{I} and r^{M} denote the optimal between-group transfers to the tagged group under immutability and mutability, respectively. It follows that

$$r^{\mathrm{M}} \leq r^{\mathrm{I}}. \tag{A13}$$

with strict inequality if the constraint (A12) binds at the immutable optimum. This proves that the optimal transfer under mutability is (weakly) lower than under immutability, completing the proof.

In the extreme case of full (two-sided) mutability, where either group may misreport, the tag loses all informational value. Then the planner must choose identical allocations for tagged and untagged individuals, implying that the optimal between-group transfer is zero.

Proof of Proposition 4

Setup: Let $n_{g,i} > 0$ denote individual consumption needs and, following Boadway & Pestieau (2006), define needs-adjusted consumption as $\tilde{c}_{g,i} = c_{g,i} - n_{g,i}$. This redefinition shifts consumption, but preserves the structure of the planner's problem unchanged.

Definition 3 (Strength of tag-needs correlation) *Let $\beta \in [0, 1]$ parameterize the strength of the correlation between the tag and needs. An increase in β, holding the pre-change between-group transfer fixed, implies*

$$\frac{\partial}{\partial \beta}\big(\mathbb{E}_t[n] - \mathbb{E}_u[n]\big) > 0.$$

Proposition 4 states: *The optimal between-group transfer is increasing in the strength of correlation of the tag with consumption needs.* When β increases, $\partial r/\partial \beta > 0$.

Proof. Starting from an optimal allocation and between-group transfer that satisfy Equation (A11), suppose we increase β as in Definition 3. By definition, the tagged group's average needs-adjusted consumption $\mathbb{E}_t[\tilde{c}]$ falls while the untagged group's average needs-adjusted

consumption $\mathbb{E}_u[\tilde{c}]$ rises. Since marginal utility is strictly decreasing in needs-adjusted consumption, the tagged group's average marginal utility of consumption increases while the untagged group's decreases, violating Equation (A11). Following the same logic as in the proof of Proposition 2, the planner must increase the transfer to the tagged group until Equation (A11) is satisfied. Hence, $\partial r/\partial\beta > 0$.

B Variable Definitions

Efficiency Concerns

Tagging incr labor supply: Indicator variable equal to one if a participant believes that tagging would encourage high-income individuals to "Work more " or "Work much more."
Tagging incr revenue: Indicator variable equal to one if the participant believes tagging would "Increase" or "Strongly increase" government revenue.
Tagging incr evasion: Indicator variable equal to one if a participant believes that, if tagging is implemented, individuals are "likely" or "very likely" to misreport their characteristics to lower their tax burden.
Tagging distorts behaviors: Indicator variable equal to one if a participant believes that, if tagging is implemented, individuals are "Likely" or "Very likely" to change their behaviors to legally lower their tax burden?
Tagging incr admin costs: Indicator variable equal to one if a participant believes tagging would "increase" or "strongly increase" the administrative costs of the government.
Tagging incr complexity: Indicator variable equal to one if a participant believes that tagging would make the process of filing taxes "more complicated" or "much more complicated."
Belief govt is inefficient: Indicator variable equal to one if a participant's belief about the cents wasted by the government is above the median belief.

Fairness Concerns

Incomes w/ tagging fair: Indicator variable equal to one if a participant believes the income distribution resulting from tagging is "Fair" or "Very fair."

Privacy Concerns

Tagging invades privacy: Indicator variable equal to one if a participant believes tagging would lead to a "moderate invasion of privacy" or "significant invasion of privacy."

Scope of Government

Govt should do more: Indicator variable equal to one if a participant believes that the "government should do more" or "government should do a lot more" to solve the country's problems.

Implementation Concerns

Tagging diff to implement: Indicator variable equal to one if a participant believes that tagging would be "difficult" or "very difficult" to implement based on legal or administrative difficulties.

High trust in govt: Indicator variable equal to one if a participant believes the federal government can be trusted "most of the time" or "just about always."

C Additional Figures

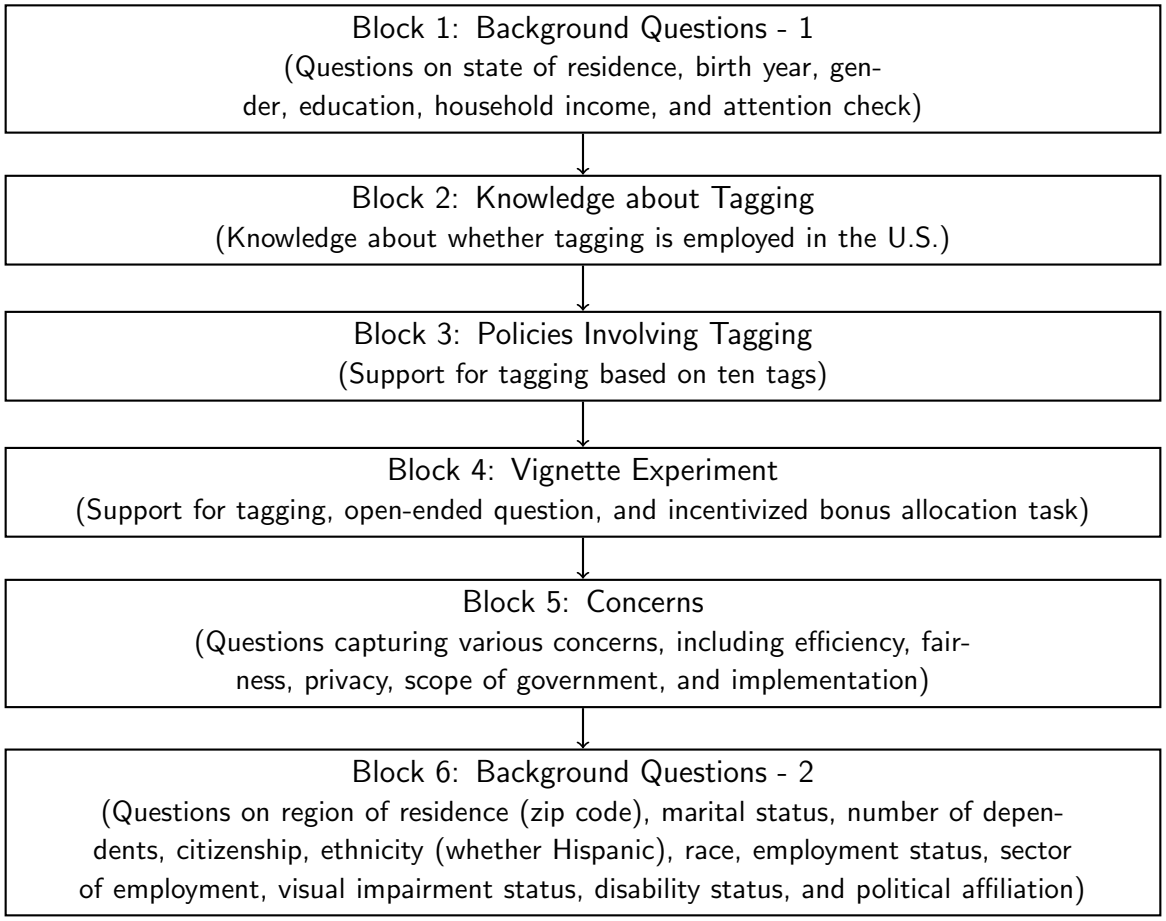


Figure A1: Overview of Experimental Design

We will now ask you about another scenario.

Suppose we divide the U.S. population into two groups that the government can identify: Group A and Group B.

Whether an individual belongs to Group A or Group B depends on a single personal characteristic over which individuals have no control. Thus, **individuals cannot influence the group to which they belong.** For example, the groups could represent individuals of different sexes or individuals with and without visual impairments.

The **average pre-tax income of individuals in Group A is lower** than the average pre-tax income of individuals in Group B. **Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income.** Thus, the incomes of the two groups are not too different, despite the lower average income in Group A.

Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs, such as housing, transportation, utilities, healthcare, and education. In society, some groups, such as those with visual impairments or dependents, have to spend more money to meet their basic needs.

Figure A2: Screenshot of Instructions in the Vignette Experiment in Treatment Immutable × Low-Corr × Uncorr-Needs

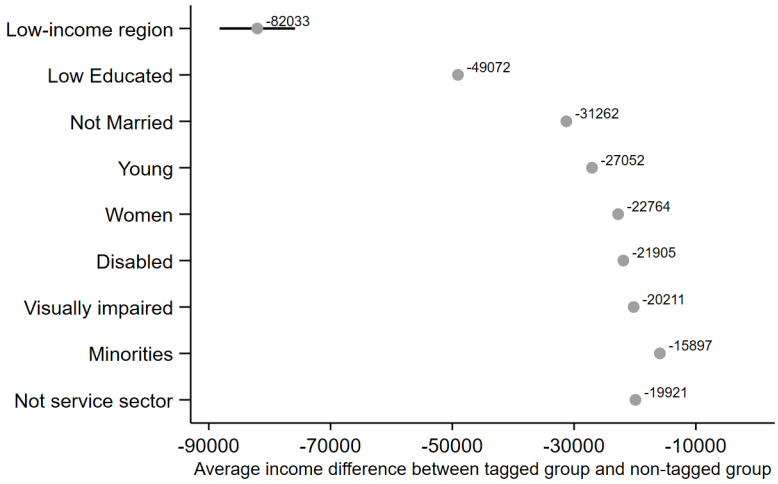


Figure A3: Average Difference in Incomes of Groups

Notes: The figure presents the average income differences across various groups. The joint distribution of income and region of residence (zip code) is based on the American Community Survey (ACS) 2022 5-year estimates, with no sample exclusions. The joint distribution of income and all other characteristics is drawn from the ACS 2023 1-year estimates (Ruggles et al. 2024). Income is measured as individual pre-tax income. The sample excludes individuals younger than 18 and those who worked fewer than 500 hours in the previous year. We do not report income differences by dependent status due to a lack of data.

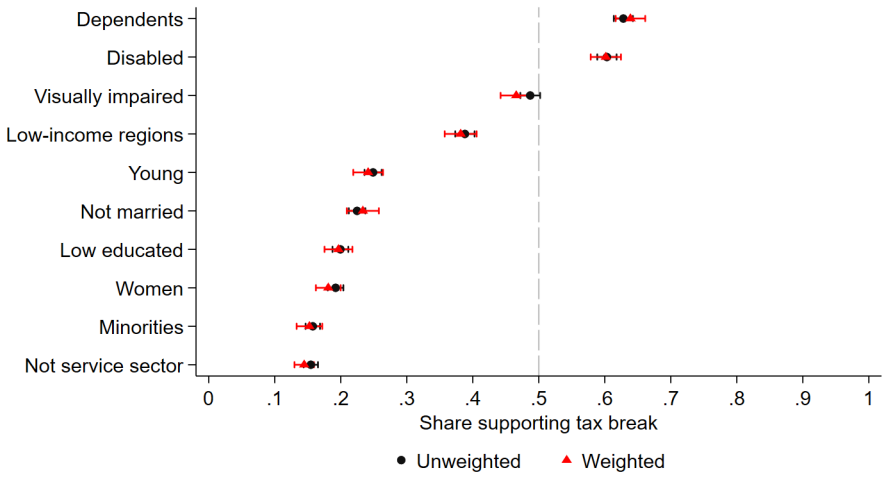


Figure A4: Share Supporting Tags, Weighted Estimates

Notes: The figure presents the share of participants who support providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the share of participants supporting a tax break for each group. A participant is defined as supporting a tax break if their responses range from "+1: Slightly support" to "+3: Strongly support." Circles (black) present unweighted estimates, while triangles (red) present weighted estimates using sampling weights. The bars represent the 95% confidence intervals around the means.

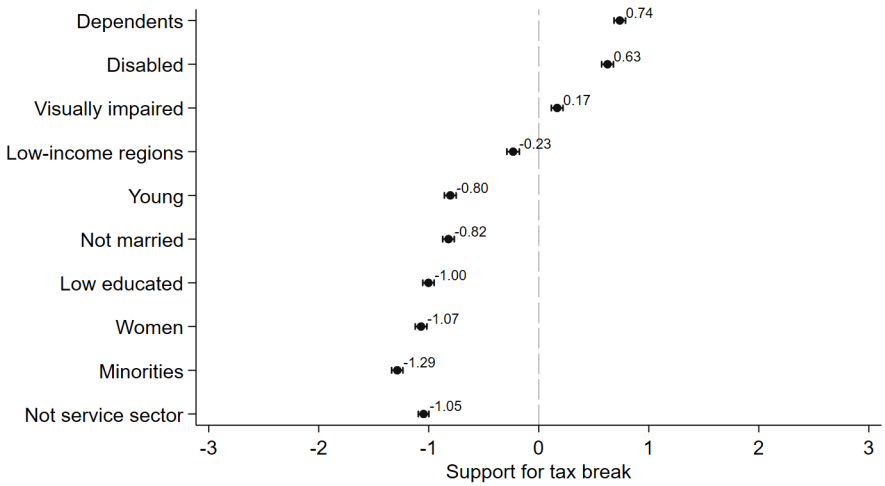


Figure A5: Average Support for Tags

Notes: The figure presents the average support for providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the support for providing a tax break for each group. Support ranges from "-3: Strongly oppose" to "+3: Strongly support." The bars represent the 95% confidence intervals around the means.

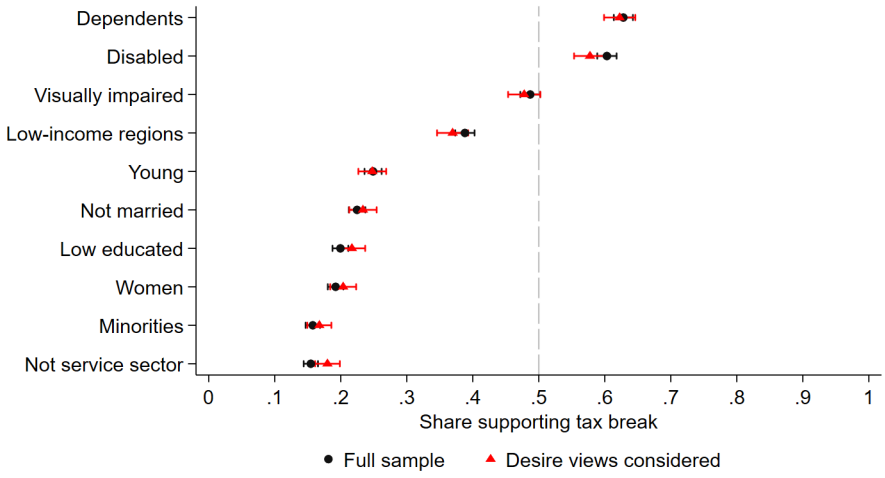


Figure A6: Share Supporting Tags, Among Strong Views to Inform Authorities

Notes: The figure presents the share of participants who support providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the share of participants supporting a tax break for each group. A participant is defined as supporting a tax break if their responses range from "+1: Slightly support" to "+3: Strongly support." Circles (black) present estimates using the full sample, while triangles (red) present estimates using only the 38% of participants with a strong desire for their views to be considered by public authorities (indicating "Very much"). The distribution of responses to this question is: "Not at all" (3%), "Very little" (5%), "Little" (13%), "Somewhat" (41%), and "Very much" (38%). The bars represent the 95% confidence intervals around the means.

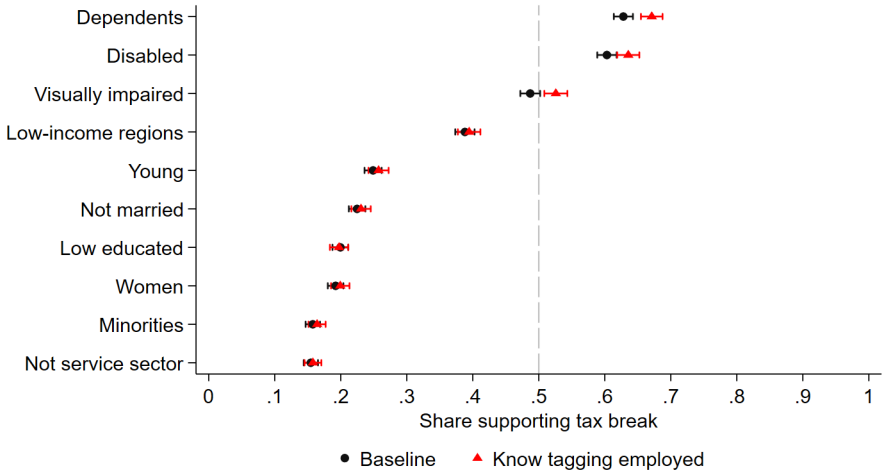


Figure A7: Share Supporting Tags, Among Knowledgeable

Notes: The figure presents the share of participants who support providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the share of participants supporting a tax break for each group. A participant is defined as supporting a tax break if their responses range from "+1: Slightly support" to "+3: Strongly support." Circles (black) present estimates using the full sample, while triangles (red) present estimates using the 74% of participants who know that tagging is employed by the tax system. The bars represent the 95% confidence intervals around the means.

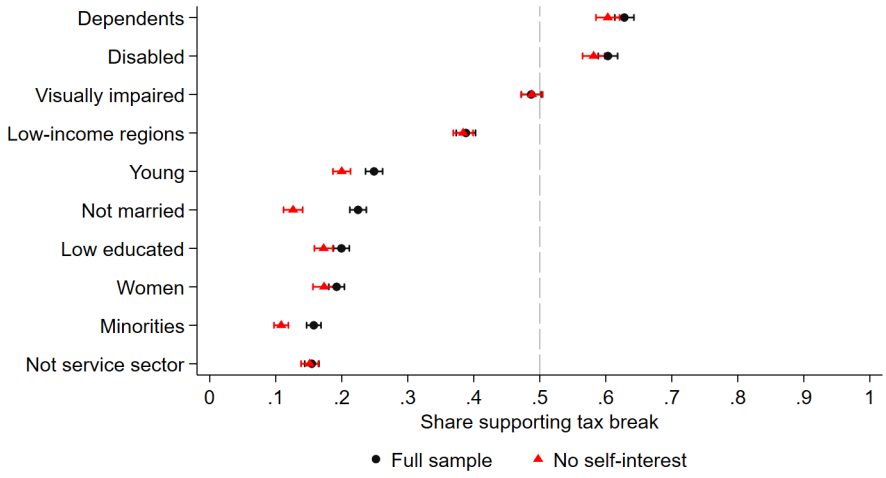


Figure A8: Share Supporting Tags, Accounting for Self-Interest

Notes: The figure presents the share of participants who support providing a tax break to various groups. The y-axis lists the groups, and the x-axis indicates the share of participants supporting a tax break for each group. A participant is defined as supporting a tax break if their responses range from "+1: Slightly support" to "+3: Strongly support." Circles (black) present estimates using the full sample, while triangles (red) present estimates excluding, for each group, participants who belong to that group (i.e., those who can benefit from the tag). The bars represent the 95% confidence intervals around the means.

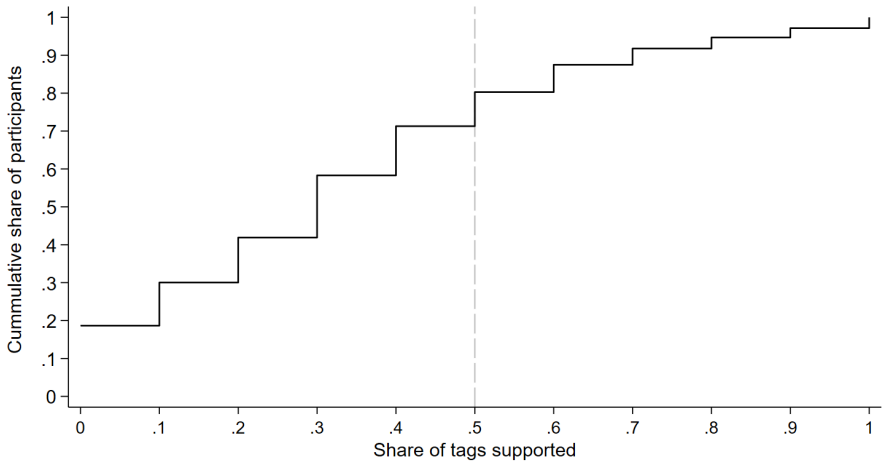


Figure A9: Distribution of Share of Tags Supported

Notes: The figure presents the cumulative distribution of the share of tags (out of ten) supported.

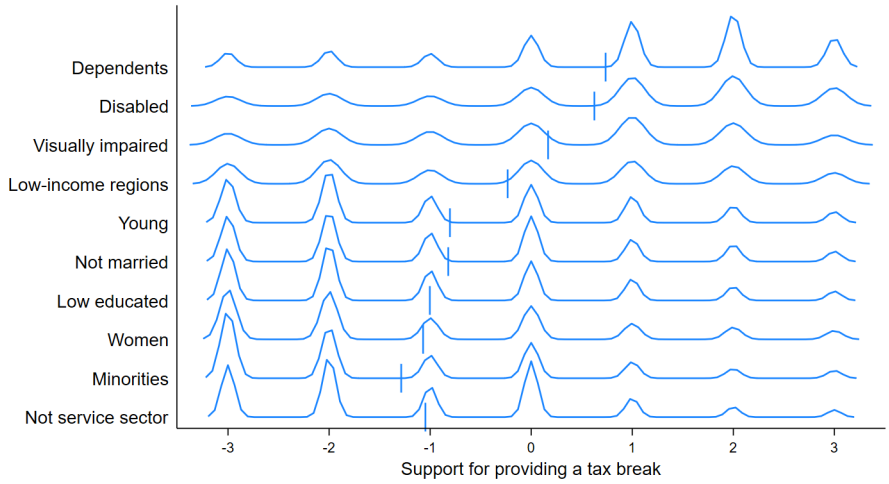


Figure A10: Distribution of Support for Tags

Notes: The figure presents the distribution of support for each tag (height of the violin) and the average support (horizontal lines) for each tag. The y-axis lists the groups, and the x-axis indicates the support for providing a tax break for each group. Support ranges from "-3: Strongly oppose" to "+3: Strongly support."

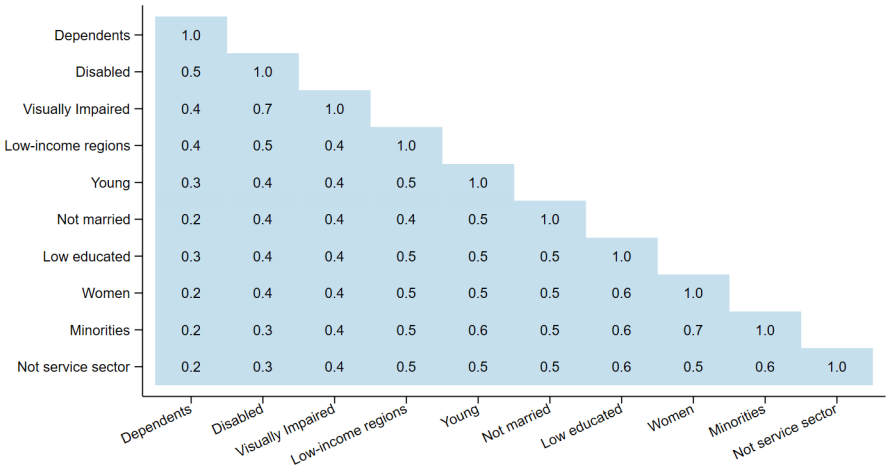


Figure A11: Pairwise Correlation in Support for Tags

Notes: The figure presents pairwise Pearson correlation coefficients of support for tags. Support ranges from "-3: Strongly oppose" to "+3: Strongly support." All correlation coefficients are statistically significant at the 5% level.

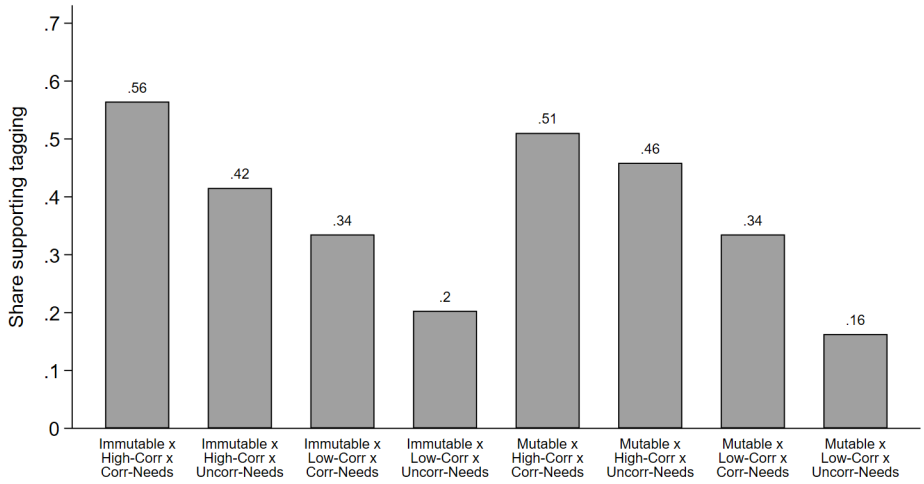
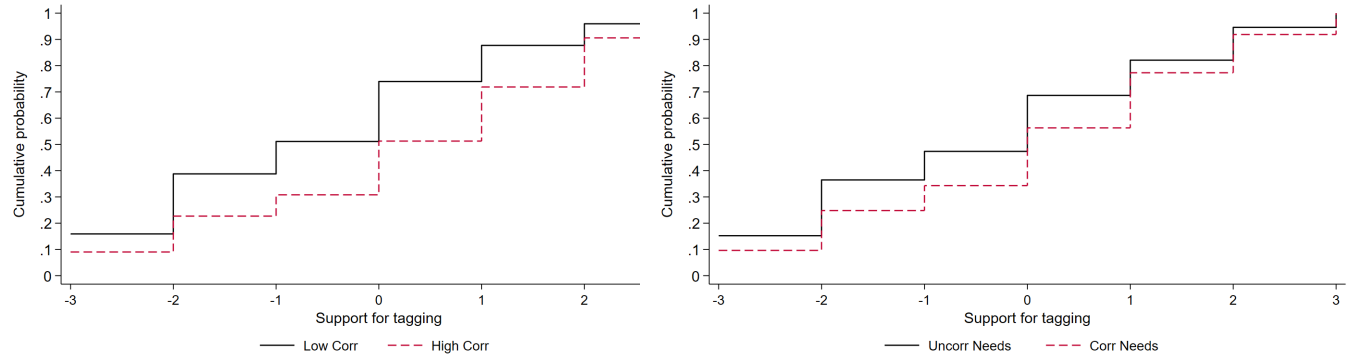


Figure A12: Share Supporting Tagging by Treatments

Notes: The figure presents the share of participants supporting tagging in the vignette experiment by each of the eight treatments.



(a) Low-Corr vs. High-Corr

(b) Uncorr-Needs vs. Corr-Needs

Figure A13: Distribution of Support for Tagging in Vignette Across Treatments

Notes: The figure presents the distribution of support for tagging in the vignette experiment across treatments. Panel (a) presents the distribution for Treatments Low-Corr and High-Corr, pooling across the other treatments. Panel (b) presents the distribution for Treatments Uncorr-Needs and Corr-Needs, pooling across the other treatments. Support ranges from "-3: Strongly oppose" to "+3: Strongly support."

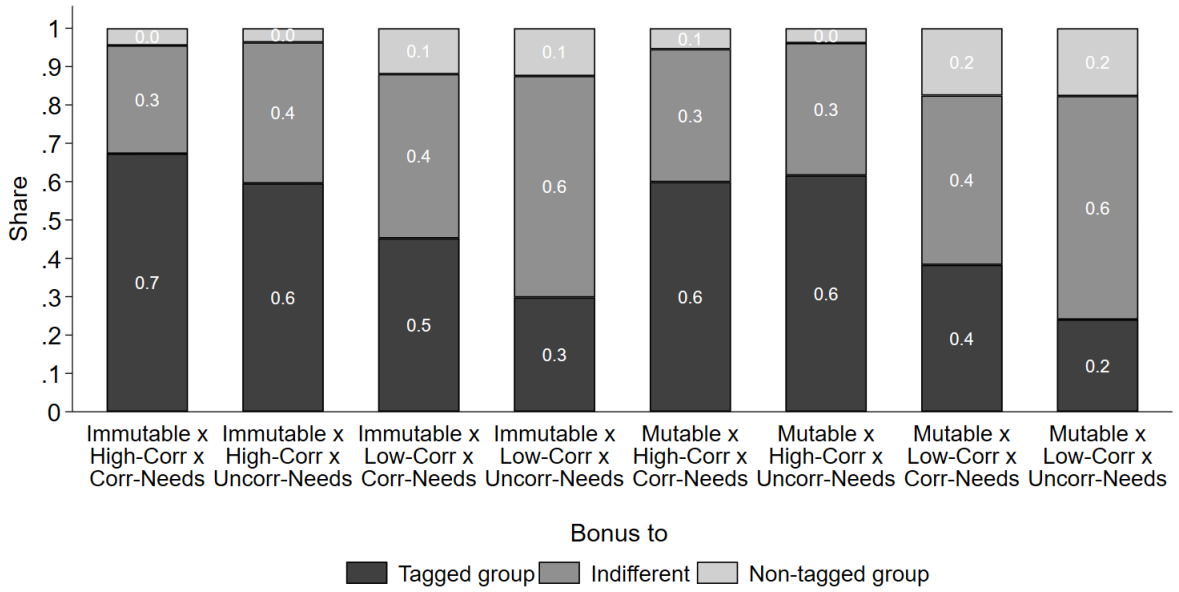


Figure A14: Distribution of Bonus Allocation

Notes: The figure presents a stacked bar graph. The y-axis presents the share of participants indicating that they would allocate the $500 bonus to the individual from the tagged group, the individual from the untagged group, or either individual. These shares are presented separately by treatment.

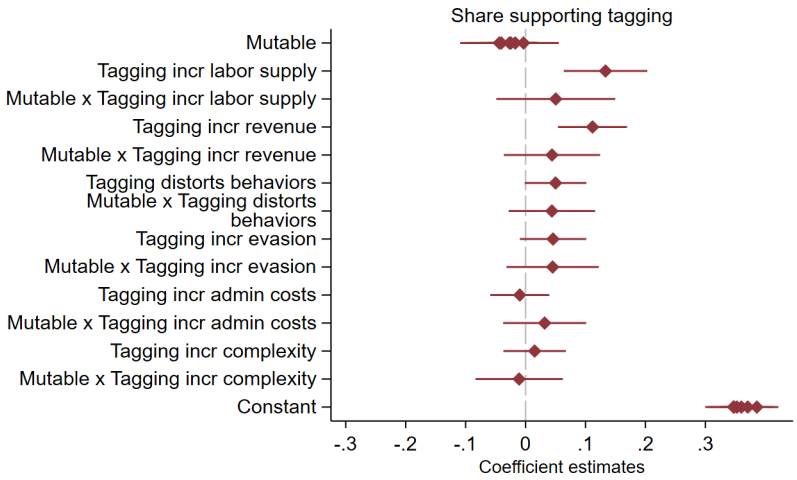


Figure A15: Heterogenous Treatment Effects of *Mutable*

Notes: The figure presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *Mutable* is a treatment indicator. The explanatory variables include variables capturing mechanisms (see Appendix Section B for details). Error bars are 95% confidence intervals based on HC3 standard errors.

D Additional Tables

Table A1: Tax Burden of Various Groups in the U.S.

Group	Tax Burden Explanation
Presence of Children	Individuals with children receive tax credits such as the Child Tax Credit and Child and Dependent Care Credit. Sources: • https://www.irs.gov/credits-deductions/individuals/child-tax-credit • https://www.irs.gov/taxtopics/tc602
Presence of Adopted Children	Individuals with adopted children receive the same benefits as those with biological children. In addition, they are eligible for the Adoption Tax Credit. Sources: • https://www.irs.gov/taxtopics/tc607
Presence of Dependents	Individuals with depends qualify for tax credits such as Child and Dependent Care Credit and Other Dependents Credit. Sources: • https://www.irs.gov/taxtopics/tc602 • https://www.irs.gov/newsroom/understanding-the-credit-for-other-dependents
Married	The tax burden of individuals who are married could be lower or higher than others. Due to the convexity of the tax schedule, individuals who are married may owe more taxes if the incomes of the spouses are far apart. However, their standard deductions are higher. Sources: • https://www.irs.gov/taxtopics/tc501 • https://www.irs.gov/newsroom/understanding-the-credit-for-other-dependents

Continued on next page

Table A1: List of Tags (continued)

Group	Tax Burden Explanation
Widow(er)	A surviving spouse may owe lower taxes in the two year period after their spouse dies due to the lower average tax rates. Sources: • https://www.irs.gov/filing/federal-income-tax-rates-and-brackets
Native Americans	Income earned directly by a Native American tribe or on tribal lands may be exempt from federal income taxes. Sources: • https://www.irs.gov/pub/irs-pdf/p5424.pdf
Seniors	Seniors can benefit from a higher standard deduction and may also be eligible for the Credit for the Elderly or the Disabled. Sources: • https://www.irs.gov/publications/p554 • https://www.irs.gov/credits-deductions/individuals/credit-for-the-elderly-or-the-disabled
Minors	The tax burden of minors for unearned income is higher due to the Kiddie Tax. Sources: • https://www.irs.gov/taxtopics/tc553
Reside in Low-Income Regions	Residents in distressed regions (Empowerment Zones) may benefit from the Empowerment Zone Employment Credit Sources: • https://www.irs.gov/forms-pubs/about-form-8844
Clergy	Clergy can deduct housing allowances and may be exempt from certain self-employment taxes. Sources: • https://www.irs.gov/taxtopics/tc417

Continued on next page

Table A1: List of Tags (continued)

Group	Tax Burden Explanation
Farmers and Fishermen	Farmers and fisherman have a lower tax burden due to rules such as income averaging. Sources: • https://www.irs.gov/forms-pubs/about-schedule-j-form-1040
Self-Employed	Self-employed individuals owe more in taxes because they owe the self-employment tax in addition to income taxes. However, their tax burden could be lower because of the various deductions available. Sources: • https://www.irs.gov/businesses/small-businesses-self-employed/self-employed-individuals-tax-center#Obligations
Students	Students are eligible for various tax credits like the American Opportunity Tax Credit and Lifetime Learning Credit. Students can also deduct education expenses and student loan interest from their income. Sources: • https://www.irs.gov/credits-deductions/individuals/aotc • https://www.irs.gov/credits-deductions/individuals/llc • https://www.irs.gov/taxtopics/tc456 • https://www.irs.gov/credits-deductions/individuals/qualified-ed-expenses
Veterans	Veterans can exclude the Department of Veterans Affairs (VA) education benefits and disability benefits from their income. In addition, they can benefit from other special tax considerations. Sources: • https://www.irs.gov/individuals/information-for-veterans

Continued on next page

Table A1: List of Tags (continued)

Group	Tax Burden Explanation
Disability Status	Disabled individuals can benefit from the Credit for the Elderly or the Disabled. In addition, they can deduct additional expenses from their income, such as impairment-related work expenses. Sources: • https://www.irs.gov/pub/irs-pdf/p907.pdf
Visually Impaired	Visually impaired individuals receive the same benefits as those who are disabled. In addition, they benefit from a higher standard deduction. Sources: • https://www.irs.gov/taxtopics/tc551
Homeowners	Homeowners can deduct their mortgage interest from their income and may be eligible for the Mortgage Interest Credit. Sources: • https://www.irs.gov/publications/p530

Table A2: Randomization Check

	Immutable				Mutable				
	Same-Needs		High-Needs		Same-Needs		High-Needs		
Variable	Low-Corr	High-Corr	Low-Corr	High-Corr	Low-Corr	High-Corr	Low-Corr	High-Corr	p-value
Female	0.55	0.56	0.53	0.57	0.57	0.53	0.53	0.57	0.74
Income: $< 30{,}000$	0.18	0.16	0.18	0.16	0.20	0.18	0.18	0.19	0.89
Income: 30–59,999	0.21	0.24	0.22	0.23	0.24	0.24	0.22	0.22	0.93
Income: 60–99,999	0.20	0.20	0.24	0.21	0.20	0.23	0.22	0.25	0.51
Income: 100–149,999	0.20	0.18	0.15	0.17	0.15	0.14	0.18	0.15	0.24
Income: $\geq 150{,}000$	0.20	0.20	0.20	0.21	0.20	0.20	0.20	0.18	0.99
Education: Up to Highschool	0.33	0.34	0.34	0.31	0.35	0.34	0.37	0.36	0.85
Education: Some college	0.16	0.16	0.17	0.22	0.17	0.19	0.18	0.19	0.37
Education: Bachelor or Associate	0.35	0.35	0.34	0.34	0.32	0.31	0.32	0.32	0.92
Education: Masters or above	0.16	0.16	0.15	0.13	0.16	0.16	0.12	0.13	0.58
Age: 18–34	0.13	0.10	0.12	0.11	0.14	0.16	0.13	0.12	0.25
Age: 25–34	0.08	0.13	0.12	0.11	0.11	0.13	0.13	0.12	0.54
Age: 35–44	0.18	0.18	0.13	0.19	0.14	0.16	0.16	0.16	0.35
Age: 45–54	0.18	0.14	0.15	0.19	0.20	0.19	0.15	0.15	0.21
Age: 55–64	0.15	0.20	0.20	0.17	0.16	0.14	0.23	0.16	0.01
Age: 65+	0.27	0.25	0.27	0.24	0.25	0.23	0.20	0.28	0.19
Region: Northeast	0.19	0.23	0.19	0.18	0.17	0.19	0.19	0.16	0.36
Region: Midwest	0.20	0.20	0.20	0.21	0.20	0.20	0.21	0.16	0.65
Region: South	0.36	0.29	0.34	0.31	0.35	0.33	0.37	0.38	0.15
Region: West	0.25	0.28	0.27	0.30	0.27	0.28	0.23	0.30	0.33
Republican	0.31	0.31	0.28	0.30	0.27	0.29	0.30	0.32	0.78

Notes: The table presents the average sample characteristics across treatments. The last column presents the p-value from an F-test from a regression of the particular characteristic on treatment indicators.

Table A3: Support for Tagging in Vignette by Treatments

	(1)	(2)	(3)	(4)	(5)
High-Corr	0.872***			0.876***	0.875***
	(0.064)			(0.063)	(0.061)
Mutable		-0.006		-0.012	-0.038
		(0.066)		(0.063)	(0.061)
Corr-Needs			0.502***	0.509***	0.519***
			(0.065)	(0.063)	(0.061)
Constant	-0.635***	-0.189***	-0.444***	-0.886***	-0.261***
	(0.044)	(0.047)	(0.047)	(0.063)	(0.091)
Observations	3012	3012	3012	3012	2983
Controls?	No	No	No	No	Yes

Notes: The table presents coefficient estimates from linear regressions. The dependent variable is participants' support for tagging in the vignette experiment, which ranges from "-3: Strongly oppose" to "+3: Strongly support." *Mutable*, *High-Corr*, and *Corr-Needs* are treatment indicators. Controls include *Republican*, *High Age*, *High Income*, *High Education*, and *Female*. *Republican* is an indicator for Republicans. *High Age* equals 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. HC3 standard errors in parentheses.

* p <0.10, ** p <0.05, *** p <0.01

Table A4: Share Supporting Tagging in Vignette by Treatments

	(1)
High-Corr	0.212***
	(0.033)
Mutable	-0.040
	(0.029)
Corr-Needs	0.132***
	(0.032)
Mutable x High-Corr	0.084*
	(0.046)
High-Corr x Corr-Needs	0.017
	(0.048)
Mutable x Corr-Needs	0.040
	(0.045)
Mutable x High-Corr x Corr-Needs	-0.138**
	(0.068)
Constant	0.203***
	(0.021)
Observations	3012

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *Mutable*, *High-Corr*, and *Corr-Needs* are treatment indicators. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A5: Heterogeneous Treatment Effects of *High-Corr*

	(1)	(2)	(3)	(4)	(5)
High-Corr	0.248***	0.219***	0.226***	0.229***	0.221***
	(0.021)	(0.029)	(0.024)	(0.025)	(0.026)
Republican	-0.106***				
	(0.023)				
High-Corr x Republican	-0.064*				
	(0.036)				
High Education		0.013			
		(0.024)			
High-Corr x High Education		0.013			
		(0.036)			
High Income			-0.083***		
			(0.023)		
High-Corr x High Income			0.001		
			(0.034)		
High Age				-0.128***	
				(0.023)	
High-Corr x High Age				-0.009	
				(0.034)	
Female					0.004
					(0.023)
High-Corr x Female					0.010
					(0.034)
Constant	0.291***	0.251***	0.298***	0.324***	0.258***
	(0.014)	(0.019)	(0.016)	(0.017)	(0.017)
High-Corr +	0.185***	0.231***	0.227***	0.220***	0.231***
High-Corr x Group Indicator	(0.029)	(0.021)	(0.025)	(0.023)	(0.023)

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *High-Corr* is a treatment indicator. *Republican* is an indicator for Republicans. *High Age* equals 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A6: Heterogeneous Treatment Effects of *Corr-Needs*

	(1)	(2)	(3)	(4)	(5)
Corr-Needs	0.146***	0.112***	0.121***	0.135***	0.133***
	(0.021)	(0.030)	(0.024)	(0.025)	(0.026)
Republican	-0.099***				
	(0.025)				
Corr-Needs x Republican	-0.072*				
	(0.037)				
High Education		0.014			
		(0.025)			
Corr-Needs x High Education		0.018			
		(0.037)			
High Income			-0.086***		
			(0.024)		
Corr-Needs x High Income			0.006		
			(0.035)		
High Age				-0.127***	
				(0.024)	
Corr-Needs x High Age				-0.020	
				(0.035)	
Female					0.020
					(0.024)
Corr-Needs x Female					-0.016
					(0.035)
Constant	0.342***	0.304***	0.353***	0.375***	0.303***
	(0.015)	(0.020)	(0.017)	(0.017)	(0.018)
Corr-Needs +	0.074**	0.130***	0.127***	0.115***	0.117***
Corr-Needs x Group Indicator	(0.030)	(0.022)	(0.025)	(0.024)	(0.024)

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *Corr-Needs* is a treatment indicator. *Republican* is an indicator for Republicans. *High Age* equals 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A7: Heterogeneous Treatment Effects of *Mutable*

	(1)	(2)	(3)	(4)	(5)
Mutable	-0.021	0.021	0.015	-0.005	0.021
	(0.021)	(0.030)	(0.025)	(0.025)	(0.026)
Republican	-0.151***				
	(0.026)				
Mutable x Republican	0.033				
	(0.037)				
High Education		0.046*			
		(0.026)			
Mutable x High Education		-0.048			
		(0.037)			
High Income			-0.052**		
			(0.025)		
Mutable x High Income			-0.062*		
			(0.035)		
High Age				-0.127***	
				(0.025)	
Mutable x High Age				-0.019	
				(0.035)	
Female					0.040
					(0.025)
Mutable x Female					-0.057
					(0.035)
Constant	0.426***	0.350***	0.406***	0.444***	0.359***
	(0.015)	(0.021)	(0.018)	(0.018)	(0.019)
Mutable +	0.012	-0.027	-0.047*	-0.024	-0.036
Mutable x Group Indicator	(0.030)	(0.022)	(0.025)	(0.024)	(0.024)

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *Mutable* is a treatment indicator. *Republican* is an indicator for Republicans. *High Age* equals 1 if a participant's age is above the median and 0 otherwise. *High Income* and *High Education* are defined similarly. *Female* is an indicator for females. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A8: Welfare Weights by Treatments

	(1)	(2)	(3)
High-Corr	0.244***		
	(0.019)		
Mutable		-0.030	
		(0.019)	
Corr-Needs			0.098***
			(0.019)
Constant	0.405***	0.551***	0.487***
	(0.014)	(0.013)	(0.014)
Observations	2724	2724	2724

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant assigns progressive welfare weights and 0 if the participant assigns equal weights. *High-Corr*, *Mutable*, and *Corr-Needs* are treatment indicators. *Progressive weights* equals 1 if a participant allocates the bonus to the individual from the tagged group and 0 if the participant is indifferent. The sample excludes participants with regressive weights. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A9: Treatment Effects Mediated by Welfare Weights

	(1)	(2)	(3)	(4)
High-Corr	0.220***	0.120***		
	(0.018)	(0.017)		
Corr-Needs			0.129***	0.087***
			(0.019)	(0.017)
Progressive weights		0.411***		0.431***
		(0.017)		(0.017)
Constant	0.273***	0.107***	0.327***	0.117***
	(0.013)	(0.012)	(0.013)	(0.012)
Observations	2724	2724	2724	2724
Share explained		0.45		0.33

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *High-Corr*, *Mutable*, and *Corr-Needs* are treatment indicators. *Progressive weights* equals 1 if a participant allocates the bonus to the individual from the tagged group and 0 if the participant is indifferent. The sample excludes participants with regressive weights. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A10: Share Supporting Tagging in Vignette by Mechanisms

	(1)
High-Corr	0.214***
	(0.016)
Mutable	-0.013
	(0.016)
Corr-Needs	0.117***
	(0.016)
Incomes w/ tagging fair	0.277***
	(0.021)
Tagging incr labor supply	0.054**
	(0.026)
Tagging incr revenue	0.059***
	(0.020)
Tagging distorts behaviors	0.029
	(0.020)
Tagging incr evasion	0.063***
	(0.022)
Tagging incr admin costs	-0.010
	(0.018)
Tagging incr complexity	0.018
	(0.020)
Belief govt is inefficient	-0.035**
	(0.017)
Govt should do more	0.068***
	(0.017)
High trust in govt	0.003
	(0.022)
Tagging diff to implement	0.015
	(0.018)
Tagging invades privacy	-0.054***
	(0.017)
High Education	0.037**
	(0.019)
Female	0.010
	(0.016)
Republican	-0.065***
	(0.018)
High Income	-0.074***
	(0.017)
High Age	-0.088***
	(0.017)
Constant	0.113***
	(0.031)
Observations	2983

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. The explanatory variables include treatment indicators, demographic controls, and variables capturing mechanisms (see Appendix Section B for details). HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A11: Decomposition of Support for Tagging in Vignette

	(1)	(2)	(3)	(4)
Fairness	0.62	0.47	0.50	0.50
Efficiency	0.18	0.14	0.23	0.21
Scope of govt	0.13	0.15	0.16	0.17
Privacy	0.05	0.04	0.05	0.03
Implementation	0.02	0.06	0.06	0.08
Dependent variable	1(Support for tagging > 0)		Support for tagging	
Explanatory variables	Binary	Continuous	Binary	Continuous
R^2	0.13	0.13	0.15	0.19

Notes: The table presents the share of R^2 in linear regressions explained by various concerns using a Shapley value decomposition (Lipovetsky 2021). In Columns (1) and (2), the dependent variable equals 1 if a participant supports tagging in the vignette and 0 otherwise, while in Columns (3) and (4), the dependent variable is support for tagging in the vignette (ranging from -3 to $+3$). In Columns (1) and (3), the explanatory variables are dummy variables (see Appendix Section B for details), while in Columns (2) and (4), the explanatory variables are standardized continuous variables. The decomposition identifies the relative contribution of concerns, each of which may include one or more variables.

Table A12: Treatment Effects Mediated by Mechanisms

	(1)	(2)	(3)	(4)
High-Corr	0.227***	0.214***		
	(0.017)	(0.016)		
Corr-Needs			0.124***	0.115***
			(0.018)	(0.016)
Incomes w/ tagging fair		0.284***		0.284***
		(0.019)		(0.020)
Tagging incr labor supply		0.051**		0.057**
		(0.024)		(0.024)
Tagging incr revenue		0.056***		0.057***
		(0.020)		(0.020)
Tagging distorts behaviors		0.030		0.027
		(0.019)		(0.019)
Tagging incr evasion		0.059***		0.067***
		(0.021)		(0.022)
Tagging decr admin costs		0.011		0.011
		(0.018)		(0.018)
Tagging incr complexity		0.022		0.020
		(0.019)		(0.020)
Belief govt is efficient		0.035**		0.045***
		(0.017)		(0.017)
Govt should do more		0.070***		0.069***
		(0.017)		(0.017)
Tagging does not invade privacy		0.054***		0.057***
		(0.017)		(0.017)
Tagging diff to implement		0.017		0.024
		(0.018)		(0.019)
High trust in govt		0.006		0.006
		(0.021)		(0.021)
Low Age		0.086***		0.093***
		(0.017)		(0.017)
Low Income		0.072***		0.075***
		(0.017)		(0.018)
High Education		0.035*		0.038**
		(0.018)		(0.019)
Female		0.008		0.011
		(0.016)		(0.017)
Not republican		0.062***		0.057***
		(0.018)		(0.019)
Constant	0.260***	-0.160***	0.313***	-0.129***
	(0.012)	(0.037)	(0.012)	(0.038)
Observations	3012	3012	3012	3012

Notes: The table presents coefficient estimates from linear regressions. The dependent variable equals 1 if a participant supports tagging in the vignette experiment and 0 otherwise. *High-Corr* and *Corr-Needs* are treatment indicators. Definitions of explanatory variables are provided in Appendix Section B, with some of these variables being reverse coded to ensure that the coefficient of all variables have a positive sign. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

Table A13: Role of Self-Interest Motives

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependents	0.076***									
	(0.018)									
Disabled		0.099***								
		(0.021)								
Visually impaired			-0.006							
			(0.025)							
Young				0.263***						
				(0.023)						
Low-income region					0.050					
					(0.049)					
Not married						0.187***				
						(0.015)				
Low educated							0.079***			
							(0.016)			
Women								0.034**		
								(0.014)		
Minorities									0.187***	
									(0.017)	
Not service sector										0.010
										(0.014)
Constant	0.603***	0.582***	0.488***	0.200***	0.384***	0.126***	0.173***	0.173***	0.108***	0.152***
	(0.011)	(0.010)	(0.010)	(0.008)	(0.009)	(0.009)	(0.008)	(0.010)	(0.007)	(0.008)
Observations	3012	3012	3012	3012	2971	3012	3012	3012	3012	3012
Share benefit	0.33	0.22	0.17	0.19	0.04	0.53	0.34	0.55	0.26	0.33

Notes: The table presents coefficient estimates from linear regressions. The explanatory variable in each column takes a value of 1 if a participant belongs to Group i and 0 otherwise. The dependent variable in each column takes a value of 1 if a participant supports a tax break for group i and 0 otherwise. *Dependents* indicates whether a participant has one or more dependents. *Disabled* indicates whether a participant or someone in their household is disabled. *Visually impaired* indicates whether a participant or someone in their household is visually impaired. *Young* indicates whether a participant is aged 18-30. *Low-income region* indicates whether a participant resides in a zip code in the bottom 1% of average earnings. *Low educated* indicates whether a participant's highest education is a high school diploma or less. *Women* indicates whether a participant is female. *Minorities* indicates whether a participant is non-white or Hispanic. *Not service sector* indicates whether a participant currently works or has ever worked in the service sector (sectors 4000 to 9290 in the 2022 census industry classifications). The last row indicates the share of participants in Group i who would benefit from the tag. HC3 standard errors in parentheses.
* p <0.10, ** p <0.05, *** p <0.01

E Additional Analysis

E.1 Correlation between Support for Various Tags

To better understand the underlying correlation structure of participants' support for tags, we conduct a Principal Component Analysis (PCA). PCA aims to uncover the underlying correlation structure among variables by identifying m components that best explain the variation in n variables, where $n > m$. These components are linear combinations of the original variables. The first component captures the largest share of variation in the data. The second component captures the largest share of the remaining variance, subject to being orthogonal to the first component. This pattern continues for subsequent components.

We conduct the PCA on participants' support for ten tags. We use three decision rules to determine how many components to retain, each of which suggests retaining two components. First, following Chapman et al. (2023), we use parallel analysis. This method generates multiple random datasets similar to the original and computes the average eigenvalues of their correlation matrices. A component is retained if its eigenvalue in the original dataset is higher than the average eigenvalue from the random datasets (Appendix Figure A16). Second, we observe an "elbow" in the eigenvalues at the second component in the scree plot (Figure A16), indicating that the marginal contribution of the third component is low. Finally, we use the Kaiser rule to retain components with eigenvalues above one, which suggests retaining only the first two components.

Table A14 presents the factor loadings of the first two principal components. The first component accounts for 50% of the total variation in the data and has a positive loading of similar magnitude for each tag. This suggests that it represents a general tendency to support or oppose tagging as a policy approach, independent of the specific tag. The positive correlations across all pairs of tags (see in Appendix Figure A11) further support this interpretation. The second component explains 13% of the total variation in the data. It has a positive loading on tags with high support (Dependents, Disabled, and Visually impaired), negative loading on tags with low support (Young, Not married, Low educated, Women, Minorities, and Not service sector), and a near-zero loading on the tag with moderate support (Low-income regions). Thus, the second component captures the contrast in support between tags with high and low levels of support.

To summarize, about half the total variation in the data appears to be explained by factors influencing individuals' general tendency to support or oppose tagging as a policy approach, regardless of the specific tag. About 13% of the variation in the data is likely captured by factors that influence support for some tags but not others.

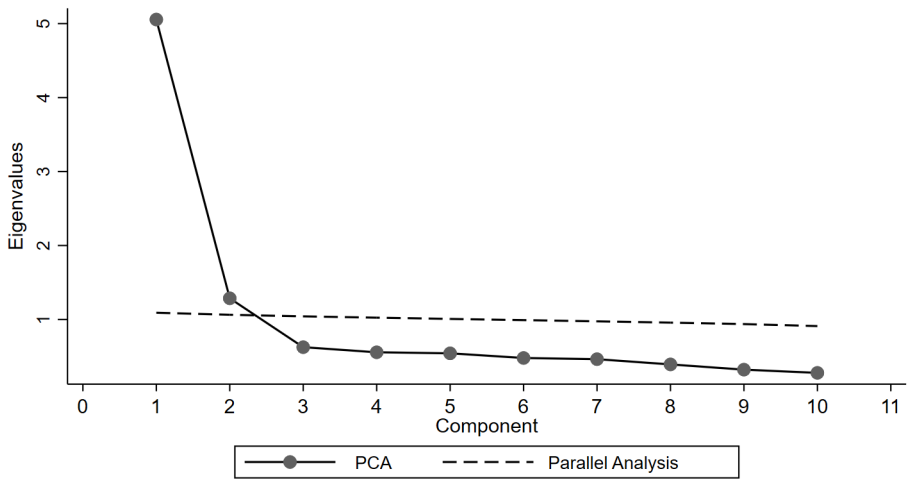


Figure A16: Scree Plot Following Principal Component Analysis (PCA)

Notes: The figure presents a scree plot from a Principal Component Analysis (PCA) of participants' support for tags, ranging from "-3: Strongly oppose" to "+3: Strongly support." The x-axis presents components, and the y-axis their eigenvalues. The solid line shows eigenvalues from the original data; the dashed line shows average eigenvalues from simulated random datasets.

Table A14: Factor Loadings from PCA

	Component	
	1	2
Dependents	0.22	0.52
Disabled	0.29	0.51
Visually impaired	0.31	0.43
Low-income regions	0.33	0.06
Young	0.33	-0.14
Not married	0.31	-0.16
Low educated	0.35	-0.22
Women	0.34	-0.24
Minorities	0.34	-0.26
Not service sector	0.33	-0.23
Share explained	0.50	0.13

Notes: The table presents factor loadings from the first two components of a Principal Component Analysis (PCA) of participants' support for tags, ranging from "-3: Strongly oppose" to "+3: Strongly support." The last row reports the share of variance explained by each component.

E.2 Open-Ended Analysis

This section describes how we analyzed participants' open-ended responses, in which they explained their support for tagging in the vignette experiment. The analysis closely follows the procedure of Bartling & Srinivasan (2025). We used OpenAI's GPT-4o model via the OpenAI API to conduct the analysis. We set the temperature parameter to 0 to obtain deterministic responses.

E.2.1 Procedures

The analysis was conducted in three sequential stages. First, we prompted the model to extract the primary reason expressed in each participant's response, cluster similar reasons based on shared underlying concerns, and tabulate the frequency of each cluster. To account for stochastic variation in model outputs, we instructed the model to repeat this procedure ten times on the same batch of responses. Second, we prompted the model to consolidate the resulting lists of reasons within each batch into a single list by combining closely related reasons and averaging their frequencies across runs. Finally, the consolidated summaries from each batch of responses were merged into a single list. In this step, the model again combined semantically similar reasons and computed the overall frequencies. The final output presents the consolidated list of reasons, frequency of occurrence of each reason, and a concise explanation.

To provide sufficient context for interpreting participants' responses, we included the following explanations in the prompt in the first step to classify responses.

Context
This prompt is about analyzing text responses from individuals who participated in a study.
Scenario
- *Participants are presented with a scenario involving **two groups** that divide the U.S. population: **Group A** and **Group B**.*
- *The government can identify which group each individual belongs to.*

Study Design
- *Participants are randomly assigned to one of **eight treatments** in a **$2 \times 2 \times 2$ between-subjects design**.*
- *Treatments vary across three dimensions:*
- ***1. Immutable vs. Mutable***
    - ***Immutable Treatment** instruction: "Whether an individual belongs to Group A or Group B depends on a single personal characteristic over which individuals*

have no control. Thus, individuals cannot influence the group to which they belong. For example, the groups could represent individuals of different sexes or individuals with and without visual impairments."

- ***Mutable Treatment** Instruction: "Whether an individual belongs to Group A or Group B depends on a single personal characteristic over which individuals have some control. Thus, individuals might be able to influence the group to which they belong. For example, the groups could represent individuals with different occupations or individuals with and without children."*

- ***2. Low-Corr vs. High-Corr (correlation between tag and income/ability)***
  - ***Common statement (both treatments):** instruction: "The average pre-tax income of individuals in Group A is lower than the average pre-tax income of individuals in Group B."*
  - ***Low-Corr Treatment** instruction: "Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income. Thus, the incomes of the two groups are not too different, despite the lower average income in Group A."*
  - ***High-Corr Treatment** instruction: "Very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income. Thus, the incomes of the two groups are very different, as reflected by the lower incomes for most individuals in Group A."*
- ***3. Uncorr-Needs vs. Corr-Needs (correlation between tag and consumption needs)***
  - ***Uncorr-Needs Treatment** instruction: "Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs, such as housing, transportation, utilities, healthcare, and education. In society, some groups, such as those with visual impairments or dependents, have to spend more money to meet their basic needs."*
  - ***Corr-Needs Treatment** instruction: "Individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs, such as housing, transportation, utilities, healthcare, and education. In society, some groups, such as those with visual impairments or dependents, have to spend more money to meet their basic needs."*

Policy Support Measure

- *Participants are asked to consider a **policy proposal**:*
  - *Instruction: "Individuals in Group A would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual*

in Group B."

- *Participants indicate their support on a **7-point scale**:*
  - *Scale: -3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; $+1$: Slightly support; $+2$: Support; $+3$: Strongly support.*

To obtain participants' reasons in Step 1, we combined the above prompt providing context with the following prompt:

Task Instructions:
*- Identify **participants' reasons**.*
*- **Identify the one most important reason in each response**.*
- Combine similar reasons based on the underlying concerns.
*- Calculate the **percentage of occurrence** for each combined reason.*
*- **The sum of the percentages of occurrences of all reasons should not add up to more than 100 percent**.*
*- Present your findings as a numbered list of combined reasons, each with its **percentage of occurrences** and a **concise explanation**.*
- Do not include any additional explanations or summaries.
Below are the text responses:

To obtain participants' consolidated reasons in Step 2, we use the following prompt:

Below are outputs from different runs of the same analysis. Each output contains a list of participants reasons in the study, their respective percentage of occurrence, and a concise summary.
Task Instructions:
*- Combine similar reasons from across the runs into a **consolidated list**.*
*- **The sum of the percentages of occurrences of all reasons in the consolidated list should not add up to more than 100 percent**.*
*- Present your findings in the following format: **Reasons (average percentage of occurrences): concise explanation**.*
- Order the consolidated list of reasons by the average percentage of occurrences.
- Do not include any additional explanations or summaries.

To obtain participants' consolidated reasons in Step 3, we use the following prompt:

Below are outputs from different batches. Each output contains a list of participants reasons in the study, their respective percentage of occurrence, and a concise summary.
Task Instructions:
*- Combine similar reasons from across the batches into a **consolidated list**.*
*- Present your findings in the following format: **Reasons (average percentage of occurrences): concise explanation**.*
- Order the consolidated list of reasons by the average percentage of occurrences.
- Do not include any additional explanations or summaries.

E.2.2 Reasons of Participant Who Do Not Support Tagging

Reasons Provided by the Model

1. **Taxes Based on Income Only (40%)**: Participants believe taxes should be based solely on an individual's income, ensuring those with the same income pay the same taxes, regardless of group membership.
2. **Fairness and Equality Concerns (16.5%)**: Participants emphasize that everyone should be treated equally under the tax system, arguing against different tax rates for individuals with the same income based on group membership.
3. **Complexity and Potential for Manipulation (11.25%)**: Concerns are raised about the complexity of the proposed system and the potential for individuals to manipulate group membership to pay less in taxes.
4. **Need for More Information (9.75%)**: Participants express a need for more information to make informed decisions, citing a lack of details about the policy's implications.
5. **Opposition to Group-Based Taxation (6.75%)**: Participants oppose the idea of group-based taxation, arguing it could create unnecessary divisions and conflicts within society.
6. **Support for Deductions Based on Needs (7.25%)**: Participants support tax deductions or credits for individuals with specific needs, such as disabilities or dependents, rather than group-based tax differences.
7. **Support for Flat Tax (7%)**: Some participants support a flat tax system, where everyone pays the same percentage of their income in taxes, regardless of income level or group membership.
8. **Neutral or Indifferent Responses (5%)**: Participants express neutrality or indifference, indicating they neither support nor oppose the proposal without further context.

Our Categorization of Reasons

1. **Fairness (horizontal equity:** 1. Taxes Based on Income Only (40%), 2. Fairness and Equality Concerns (16.5%), 6. Support for Deductions Based on Needs (7.25%), 7. Support for Flat Tax (7%)
2. **Efficiency:** 3. Complexity and Potential for Manipulation (11.25%)
3. **Other reasons:** 4. Need for More Information (9.75%), 5. Opposition to Group-Based Taxation (6.75%), 8. Neutral or Indifferent Responses (5%)

E.2.3 Reasons of Participant Who Support Tagging

Reasons Provided by the Model

1. **Lower Income Should Pay Less Taxes (41.67%)**: Participants believe that individuals with lower income should pay less in taxes, ensuring that taxes are proportional to income and not overburdening those with lower earnings.
2. **Higher Expenses for Basic Needs (26.67%)**: Many participants argue that individuals with higher expenses for basic needs should receive tax breaks to help manage their financial burdens.
3. **Fairness and Equity (15%)**: Some participants emphasize the importance of fairness and equity, suggesting that tax policies should aim to reduce income inequality and provide support to those who are disadvantaged.
4. **Support for Disadvantaged Groups (10%)**: Participants express support for disadvantaged groups, advocating for tax breaks to help them manage their financial burdens more effectively.
5. **Progressive Taxation Principles (5%)**: A few participants support the policy based on progressive taxation principles, where lower-income groups receive tax relief to promote economic equity.

Our Categorization of Reasons

1. **Fairness (vertical equity):** 1. Lower Income Should Pay Less Taxes (41.67%), 2. Higher Expenses for Basic Needs (26.67%), 3. Fairness and Equity (15%), 4. Support for Disadvantaged Groups (10%), 5. Progressive Taxation Principles (5%)

E.2.4 Reasons of Participant in Treatment Low-Corr

Reasons Provided by the Model

1. **Fairness and Equality in Taxation (33.3%)**: Participants believe taxes should be based solely on individual income, advocating for equal taxation for individuals with the same income, regardless of group characteristics.
2. **Income-Based Taxation (25%)**: Participants support taxes determined by individual income levels, with higher earners paying more and lower earners paying less, independent of group classification.
3. **Consideration of Basic Needs and Expenses (16.7%)**: Participants support tax breaks for individuals with higher expenses for basic needs, justifying lower taxes to help manage financial burdens.
4. **Fairness and Equity Concerns (12.3%)**: Participants express concerns about fairness, emphasizing that tax policies should not favor one group over another without a valid reason.
5. **Complexity and Confusion (10%)**: Participants find the tax proposal complex or confusing, suggesting a need for a simpler tax system.
6. **Support for Lower-Income Individuals (8.3%)**: Participants support tax relief for lower-income individuals, acknowledging their economic challenges and promoting a fairer tax system.
7. **Need for More Information (9%)**: Participants indicate a need for more information about the policy and group classifications before making an informed decision.
8. **Opposition to Group-Based Taxation (5%)**: Participants oppose group-based taxation, arguing it could lead to unfair treatment and social division.
9. **Support for Special Circumstances (5%)**: Participants support tax breaks for individuals with special circumstances, such as disabilities or dependents, to help meet financial challenges.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Fairness and Equality in Taxation (33.3%), 2. Income-Based Taxation (25%), 4. Fairness and Equity Concerns (12.3%)
2. **Fairness (vertical equity):** 3. Consideration of Basic Needs and Expenses (16.7%), 6. Support for Lower-Income Individuals (8.3%), 9. Support for Special Circumstances (5%)
3. **Efficiency:** 5. Complexity and Confusion (10%)

4. **Other:** 6. Need for More Information (9%), 7. Opposition to Group-Based Taxation (5%)

E.2.5 Reasons of Participant in Treatment High-Corr

Reasons Provided by the Model

1. **Income-Based Taxation (31.88%)**: Many participants believe that taxes should be based solely on income, with those earning less paying less in taxes, reflecting a preference for a progressive tax system.
2. **Support for Lower-Income Individuals/Groups (19.38%)**: Participants support the policy because it provides necessary relief to lower-income individuals or groups, helping them meet basic needs and improve their financial situation.
3. **Fairness and Equality (20.75%)**: Participants argue that individuals with the same income should pay the same amount in taxes, emphasizing fairness and equality in the tax system.
4. **Concerns About Fairness and Equity (10%)**: Participants express concerns about the fairness and equity of the proposed policy, questioning the justice of taxing individuals differently based on group classification rather than income alone.
5. **Need for More Information (7.5%)**: Some participants feel they need more information to make an informed decision, indicating uncertainty about the policy's implications.
6. **Support for Tax Relief Due to Higher Expenses (7%)**: Participants support tax relief for individuals who have higher expenses for basic needs, suggesting that those who spend more on necessities should receive some form of tax break.
7. **Concerns About Manipulation and Misuse (5%)**: Participants worry that the policy could be exploited, leading to fraud or unintended consequences, and caution against creating loopholes that could be manipulated.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Income-Based Taxation (31.88%), 3. Fairness and Equality (20.75%), 4. Concerns About Fairness and Equity (10%)
2. **Fairness (vertical equity):** 2. Support for Lower-Income Individuals/Groups (19.38%), 6. Support for Tax Relief Due to Higher Expenses (7%)
3. **Efficiency:** 7. Concerns About Manipulation and Misuse (5%)
4. **Other:** 5. Need for More Information (7.5%),

E.2.6 Reasons of Participant in Treatment Immutable

Reasons Provided by the Model

1. **Income-Based Taxation (35%)**: Many participants believe that taxes should be based solely on income, with higher earners paying more taxes, regardless of group characteristics.
2. **Lower Income, Lower Taxes (24.67%)**: Participants support lower taxes for individuals with lower income and higher expenses for basic needs, justifying a tax break to help them manage their financial burdens better.
3. **Fairness and Equality (21%)**: Participants argue that everyone should pay the same tax rate if they earn the same income, emphasizing fairness and equality in the tax system without considering group characteristics.
4. **Opposition to Group-Based Taxation (9.33%)**: Participants oppose the idea of determining taxes based on group characteristics, viewing it as discriminatory or unfair, and prefer a system based on individual circumstances.
5. **Complexity and Uncertainty (10.33%)**: Participants express confusion or uncertainty about the policy, citing the complexity of the tax system and the lack of clear criteria for group assignment.
6. **Concerns About Complexity and Fairness (7%)**: Participants express concerns about the complexity and fairness of implementing a tax system based on group characteristics, fearing arbitrary or unfair taxation.
7. **Support for Needs-Based Adjustments (7.67%)**: Participants support tax adjustments based on specific needs, such as disabilities or higher living costs, rather than group membership.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Income-Based Taxation (35%), 3. Fairness and Equality (21%), 4. Opposition to Group-Based Taxation (9.33%)
2. **Fairness (vertical equity):** 2. Lower Income, Lower Taxes (24.67%), 7. Support for Needs-Based Adjustments (7.67%)
3. **Efficiency:** 5. Complexity and Uncertainty (10.33%), 6. Concerns About Complexity and Fairness (7%)

E.2.7 Reasons of Participant in Treatment Mutable

1. **Income-Based Taxation (29.7%)**: Participants believe taxes should be based solely on individual income, not on group characteristics, ensuring individuals with the same income pay the same taxes.
2. **Support for Lower-Income Individuals (26.4%)**: Participants support lower taxes for individuals with lower incomes due to their limited financial capacity and greater challenges in meeting basic needs.
3. **Consideration of Basic Needs and Expenses (12.8%)**: Participants argue for tax breaks for individuals with higher expenses for basic needs, to alleviate their financial burden.
4. **Fairness and Equity Concerns (12.7%)**: Participants express concerns about fairness and equity, emphasizing the importance of treating everyone equally in the tax system.
5. **Complexity and Confusion (8.3%)**: Participants find the policy proposal complex or confusing, expressing a desire for a simpler tax system.
6. **Need for More Information (5.7%)**: Participants feel they need more information about the groups and the proposal to make an informed decision.
7. **Support for Progressive Taxation (5.3%)**: Participants support progressive taxation, where higher-income individuals pay more, promoting economic equity.
8. **Opposition to Group-Based Taxation (5%)**: Participants oppose group-based taxation, arguing it could lead to discrimination or unfair treatment based on personal characteristics.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Income-Based Taxation (29.7%), 4. Fairness and Equity Concerns (12.7%), 8. Opposition to Group-Based Taxation (5%)
2. **Fairness (vertical equity):** 2. Support for Lower-Income Individuals (26.4%), 3. Consideration of Basic Needs and Expenses (12.8%), 7. Support for Progressive Taxation (5.3%),
3. **Efficiency:** 5. Complexity and Confusion (8.3%)
4. **Other:** 6. Need for More Information (5.7%)

E.2.8 Reasons of Participant in Treatment Uncorr-Needs

Reasons Provided by the Model

1. **Income-Based Taxation (34.73%)**: Participants believe taxes should be based on individual income, ensuring those with higher incomes pay more, while those with lower incomes pay less, promoting fairness and proportionality.
2. **Support for Lower-Income Individuals (18.13%)**: Participants support tax relief for lower-income individuals to help meet basic needs and reduce income inequality.
3. **Fairness and Equality (20.17%)**: Concerns about fairness and equality are expressed, with participants advocating for equal tax treatment based on income, not arbitrary group characteristics.
4. **Complexity and Need for More Information (11.33%)**: Participants express concerns about the complexity of the tax system and the need for more information to make informed decisions.
5. **Opposition to Group-Based Taxation (6.13%)**: Participants oppose group-based taxation, arguing it could lead to unfair advantages or disadvantages and should be based on individual circumstances.
6. **Progressive Taxation Principles (6%)**: Some participants support progressive taxation, where higher incomes pay a larger share, to promote social equity.
7. **Support for Specific Needs (7.05%)**: Participants support tax policies that consider special circumstances, such as disabilities or dependents, which may require additional financial support.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Income-Based Taxation (34.73%), 3. Fairness and Equality (20.17%), 6. Progressive Taxation Principles (6%)
2. **Fairness (vertical equity):** 2. Support for Lower-Income Individuals (18.13%), 7. Support for Specific Needs (7.05%)
3. **Efficiency:** 4. Complexity and Need for More Information (11.33%)
4. **Other:** 5. Opposition to Group-Based Taxation (6.13%)

E.2.9 Reasons of Participant in Treatment Corr-Needs

Reasons Provided by the Model

1. **Income-Based Taxation (31.19%)**: Participants believe that taxes should be based solely on income, ensuring that individuals with the same income pay the same taxes, regardless of group affiliation or living expenses.
2. **Higher Expenses Justify Lower Taxes (25%)**: Many participants argue that individuals with higher expenses for basic needs should pay lower taxes to help manage their financial burdens.
3. **Fairness and Equality in Taxation (21.25%)**: Participants emphasize that everyone should pay the same tax rate if they earn the same income, advocating for fairness and equality in taxation.
4. **Complexity and Implementation Concerns (9.94%)**: Participants express concerns about the complexity of the tax policy and the potential for fraud or manipulation, indicating a need for more information to make an informed decision.
5. **Support for Disadvantaged Groups (8.75%)**: Participants support tax relief for disadvantaged groups, such as those with disabilities or higher living costs, to promote equity.
6. **Opposition to Group-Based Taxation (7.5%)**: Participants oppose determining taxes based on group characteristics, arguing it could lead to discrimination or unfair advantages.
7. **Moral and Ethical Considerations (5%)**: Participants mention moral and ethical reasons for supporting or opposing the policy, such as helping those in need or concerns about social engineering.
8. **Neutral or Indifferent Stance (5%)**: A small group of participants neither supports nor opposes the policy, expressing neutrality or indifference towards the proposal.

Our Categorization of Reasons

1. **Fairness (horizontal equity):** 1. Income-Based Taxation (31.19%), 3. Fairness and Equality in Taxation (21.25%)
2. **Fairness (vertical equity):** 2. Higher Expenses Justify Lower Taxes (25%), 5. Support for Disadvantaged Groups (8.75%)
3. **Efficiency:** 4. Complexity and Implementation Concerns (9.94%)
4. **Other:** 6. Opposition to Group-Based Taxation (7.5%), 7. Moral and Ethical Considerations (5%), 8. Neutral or Indifferent Stance (5%)

F Experimental Instructions

Bold text, underlining, italics, and tables appear as in the original screens.

Introduction
Welcome! This study is conducted by UniDistance Suisse and the University of Zurich. Our goal is to understand American residents' views on various topics. By completing this survey carefully, you are helping us to understand these views.

Time Required
This study will take around **14 minutes**.

Compensation
You will receive a payment for completing the study. The study includes attention checks to ensure that you carefully read the questions before answering them. If you fail the attention checks, you will not be able to complete the study and will not receive any payment.

Public outreach and confidentiality
The results of this study will be disseminated to the public. In addition, we plan to reach out to politicians to make them aware of our results. Strict confidentiality is guaranteed. Your answers will remain completely anonymous.

Voluntary participation
Participation is voluntary. You can withdraw from the study at any time.

Questions about the survey
If you have questions about this study or your rights, please get in touch with us at krishna.srinivasan@unidistance.ch

Consent
I have read the above information, and I am willing to participate.

[Yes; No]

Screen Break

[Screen shown if participant does not provide consent]

End of survey

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

Please click the next button to finish the study. You will be automatically redirected.

[**Block 1: Background Questions - 1**]

In which state do you currently reside?

[Alabama; ...; Wyoming; I do not reside in the U.S.]

In which year were you born?

[2006; ...; 1929 or earlier]

What is your gender?

[Female; Male; Non-binary; Prefer not to say]

Which category best describes your highest level of education?

[Primary education or less; Some high school; High school degree/GED; Some college; 2-year college degree; 4-year college degree; Master's degree; Doctoral degree; Professional degree (e.g., JD, MD, MBA)]

What was the TOTAL income of your **household**, before taxes, last year?

[$0-$9,999; $10,000-$19,999; $20,000-$29,999; $30,000-$39,999; $40,000-$49,999; $50,000-$59,999; $60,000-$69,999; $70,000-$79,999; $80,000-$89,999; $90,000-$99,999; $100,000-$109,999; $110,000-$119,999; $120,000-$129,999; $130,000-$139,999; $140,000-$149,999; $150,000-$159,999; $160,000-$169,999; $170,000-$179,999; $180,000-$189,999; $190,000-$199,999; $200,000-$249,999; $250,000-$299,999; $300,000 and over; Prefer not to say]

Screen Break

[Screen shown if quotas are full]

End of survey

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

Thank you for your time.

Please click the next button to finish the study. You will be automatically redirected.

Screen Break

The question below is related to the following problem. In surveys like ours, there are sometimes a few people who do not carefully read the questions they are asked and just quickly click through the survey. These random answers are problematic because they compromise the results of the studies. In order to show that you read our questions carefully, please answer "Not at all interested" below:

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

Screen Break

[Screen shown if participant failed the attention check]

End of survey

Unfortunately, you failed the attention check.

For this reason, you cannot continue the study and will not receive a payment.

Please click the next button to finish the study. You will be automatically redirected.

[Screen shown if participant passed the attention check]

In the following questions, we will explore your views on tax and transfer policies in the United States.

The anonymized results of this study will be shared with the public. In addition, we plan to reach out to politicians to make them aware of our results.

Therefore, please answer the following questions carefully and honestly.

Your answers will remain anonymous.

Screen Break

[**Block 2: Knowledge about Tagging**]

In the U.S., the federal income taxes owed by individuals depend on their income.

To your knowledge, do the taxes owed by individuals also depend on their personal characteristics, such as their marital status or their age? In other words, can two individuals

with the same income owe different amounts in taxes because of differences in their personal characteristics?

[Yes, two individuals with the same income might owe different amounts in taxes; No, two individuals with the same income cannot owe different amounts in taxes; I don't know]

Screen Break

[**Block 3: Policies Involving Tagging**]

On the following screens, we will ask you about your views on different federal income tax policies. These policies might affect different groups of individuals in different ways.

Consider the following example. Suppose that we divide the population into two groups:

Group A: Individuals with visual impairments
Group B: All other individuals

Suppose that the **average pre-tax income of individuals in Group A (with visual impairments) is lower than the average pre-tax income of individuals in Group B.** This means that individuals with visual impairments tend to earn less money than the rest of the population. **Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.**

Consider a policy proposal in which **individuals in Group A (with visual impairments) would owe slightly lower income taxes compared to individuals in Group B.** This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B. For example, a visually impaired individual with an income of $30,000 would owe slightly lower taxes than an individual who also has an income of $30,000 but is *not* visually impaired.

We are interested in knowing whether you would support such a policy.

On the following screens, we will present you with **ten such policy proposals.** In each proposal, the population is split into two different groups along a different characteristic.

There are no right or wrong answers; we are just interested in your opinion.

Screen Break

[The order of the questions is randomized across participants.]

The population is divided into two groups:

Group A: **Individuals who are NOT married**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (who are NOT married) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (who are NOT married) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B. Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals aged 18 to 30**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (aged 18 to 30) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (aged 18 to 30) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Women**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (women) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (women) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Ethnic/racial minorities**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (ethnic/racial minorities) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (ethnic/racial minorities) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals with disabilities**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (with disabilities) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (with disabilities) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals NOT working in the service sector**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (NOT working in the service sector) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (NOT working in the service sector) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals whose highest education is a high school diploma or less**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (whose highest education is a high school diploma or less) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (whose highest education is a high school diploma or less) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals with visual impairments**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (with visual impairments) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (with visual impairments) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

Imagine we divide the U.S. into 100 regions, ranked by income and poverty. Region #1 has the lowest average income and highest poverty rate in the U.S. Region #100 has the highest average income and lowest poverty rate.

Group A: **Individuals living in region #1, with the lowest average income and highest poverty rate**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (living in region #1) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (living in region #1) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

The population is divided into two groups:

Group A: **Individuals with dependents (who financially support their children or other relatives)**
Group B: All other individuals

Suppose that the average pre-tax income of individuals in Group A (with dependents) is lower than the average pre-tax income of individuals in Group B. Nevertheless, some individuals in Group A earn a high income, and some individuals in Group B earn a low income.

Policy Proposal: Individuals in Group A (with dependents) would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Screen Break

In the previous screens, you indicated your support for various policy proposals.

To what extent do you think public authorities should consider your answers when deciding on tax policies?

[Not at all; Very little; Little; Somewhat; Very much]

Screen Break

[**Block 4: Vignette Experiment**]

[Participants are randomly assigned to one of 8 treatments in a $2 \times 2 \times 2$ between-subjects design. First dimension is Immutable vs Mutable, second dimension is Low-Corr vs. High-Corr, and the third dimension is Uncorr-Needs vs. Corr-Needs.]

We will now ask you about another scenario.

Suppose we divide the U.S. population into two groups that the government can identify: Group A and Group B.

[Treatment Immutable] Whether an individual belongs to Group A or Group B depends on a single personal characteristic over which individuals have no control. Thus, **individuals cannot influence the group to which they belong.** For example, the groups could represent individuals of different sexes or individuals with and without visual impairments.

[Treatment Mutable] Whether an individual belongs to Group A or Group B depends on a single personal characteristic over which individuals have some control. Thus, **individuals might be able to influence the group to which they belong.** For example, the groups could represent individuals with different occupations or individuals with and without children.

[Treatment Low-Corr] The **average pre-tax income of individuals in Group A is lower** than the average pre-tax income of individuals in Group B. **Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income.** Thus, the incomes of the two groups are not too different, despite the lower average income in Group A.

[Treatment High-Corr] The **average pre-tax income of individuals in Group A is lower** than the average pre-tax income of individuals in Group B. **Very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income.** Thus, the incomes of the two groups are very different, as reflected by the lower incomes for most individuals in Group A.

[Treatment Uncorr-Needs] **Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs,** such as housing, transportation, utilities, healthcare, and education. In society, some groups, such as those with visual impairments or dependents, have to spend more money to meet their basic needs.

[Treatment Corr-Needs] **Individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs,** such as

housing, transportation, utilities, healthcare, and education. In society, some groups, such as those with visual impairments or dependents, have to spend more money to meet their basic needs.

Please answer the following questions to demonstrate that you have understood the instructions.

Is this statement True or False?

[Treatment Immutable] *Individuals cannot influence the group to which they belong.*

[Treatment Mutable] *Individuals might be able to influence the group to which they belong.*

[True; False]

Is this statement True or False?

[Treatment Low-Corr] *The average pre-tax income of individuals in Group A is lower. Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income.*

[Treatment High-Corr] *The average pre-tax income of individuals in Group A is lower. Very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income.*

[True; False]

Is this statement True or False?

[Treatment Uncorr-Needs] *Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs.*

[Treatment Corr-Needs] *Individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs.*

[True; False]

Screen Break

[Screen shown if participant failed one or more comprehension checks]

End of survey

Unfortunately, you did not correctly answer the comprehension questions.

For this reason, you cannot continue the study.

Please click the next button to finish the study. You will be automatically redirected.

Screen Break

Recall the following information on the two groups we just described:

- The average income of individuals in Group A is lower than the average income of individuals in Group B.
- [Treatment Low-Corr] Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income.
- [Treatment High-Corr] Very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income.
- [Treatment Uncorr-Needs] Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs.
- [Treatment Corr-Needs] Individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs.
- [Treatment Immutable] Individuals cannot influence the group to which they belong.
- [Treatment Mutable] Individuals might be able to influence the group to which they belong.

Consider the following policy proposal:

Individuals in Group A would owe slightly lower income taxes compared to individuals in Group B. This means that, among all individuals with the same income, an individual in Group A would owe less taxes than an individual in Group B.

Would you oppose or support this proposal?

[-3: Strongly oppose; -2: Oppose; -1: Slightly oppose; 0: Neither oppose nor support; +1: Slightly support; +2: Support; +3: Strongly support]

Please explain your reasoning for supporting or opposing the above policy proposal.

Please use the text box below and write as much as you like.

Screen Break

Recall the following information on the two groups we just described:

- The average income of individuals in Group A is lower than the average income of individuals in Group B.
- [Treatment Low-Corr] Nevertheless, many individuals in Group A earn a very high income, and many individuals in Group B earn a very low income.
- [Treatment High-Corr] Very few individuals in Group A earn a high income, and very few individuals in Group B earn a low income.
- [Treatment Uncorr-Needs] Individuals in Group A and Group B spend the same amount of money, on average, to meet their basic needs.
- [Treatment Corr-Needs] Individuals in Group A have to spend more money, on average, compared to individuals in Group B to meet their basic needs.
- [Treatment Immutable] Individuals cannot influence the group to which they belong.
- [Treatment Mutable] Individuals might be able to influence the group to which they belong.

We will now ask you to make a **decision that may have real monetary consequences for two other individuals.**

We will recruit two individuals from the general population of the U.S. based on the description provided above. One individual corresponds to the description provided for Group A, and the other corresponds to the description provided for Group B.

Your task is to decide whether to allocate a $500 bonus to the individual from Group A, or to the individual from Group B. If you have no preference, we will randomly decide which of the two individuals receives the bonus.

At the end of the study, we will randomly select one participant and implement their decision. Thus, it is possible that we will pay the bonus according to your decision.

Please make your decision:

[I prefer giving $500 to the individual from Group A; I prefer giving $500 to the individual from Group B; I am indifferent between the two options]

Screen Break

[**Block 5: Concerns**]

In the U.S., the federal income taxes owed by individuals depend on their income. However, these taxes could also depend on individuals' personal characteristics, such as their sector of employment or disability status. As you saw in the previous screens, this could involve lower tax rates for groups with lower average incomes.

We want to understand **what considerations come to your mind when you think about whether the taxes owed by individuals should depend on their personal characteristics in addition to their incomes.**

If taxes were based on individuals' personal characteristics in addition to their income, I think that the administrative costs of the government would ...

[Strongly decrease; Decrease; Slightly decrease; Be unaffected; Slightly increase; Increase; Strongly increase]

If taxes were based on individuals' personal characteristics in addition to their income, I think that the process of filing taxes for taxpayers would be ...

[Much less complicated; Less complicated; Slightly less complicated; Unaffected; Slightly more complicated; More complicated; Much more complicated]

Out of every tax dollar that goes to the federal government, how many cents would you say are wasted?

You can enter any number between (and including) 0 and 100.

Screen Break

If taxes were based on individuals' personal characteristics in addition to their income, I think that high-income individuals would ...

[Work much less; Work less; Work slightly less; Neither work less nor more; Work slightly more; Work more; Work much more]

If taxes were based on individuals' personal characteristics in addition to their income, I think that the total revenue the government collects would ...

[Strongly decrease; Decrease; Slightly decrease; Be unaffected; Slightly increase; Increase; Strongly increase]

If taxes were based on individuals' personal characteristics in addition to their income, how likely do you think individuals would be to misreport their characteristics to lower their tax burden?

[Very unlikely; Unlikely; Somewhat unlikely; Neither unlikely nor likely; Somewhat likely; Likely; Very likely]

If taxes were based on individuals' personal characteristics in addition to their income, how likely do you think individuals would be to change their behaviors to legally lower their tax burden?

[Very unlikely; Unlikely; Somewhat unlikely; Neither unlikely nor likely; Somewhat likely; Likely; Very likely]

Screen Break

If taxes were based on individuals' personal characteristics in addition to their income, I think that the resulting distribution of individuals' after-tax incomes would be ...

[Very unfair; Unfair; Somewhat unfair; Neither unfair nor fair; Somewhat fair; Fair; Very fair]

If taxes were based on individuals' personal characteristics in addition to their income, to what extent do you think this would invade individuals' privacy?

[No invasion of privacy; Slight invasion of privacy; Moderate invasion of privacy; Significant invasion of privacy]

Some people think the government is trying to do too many things that should be left to

individuals and businesses. Others think that the government should do more to solve our country's problems. Which comes closer to your own view?

[Government is doing way too much; Government is doing too much; Government is doing a little bit too much; Government is doing just the right amount; Government should do a little more; Government should do more; Government should do a lot more]

Screen Break

Considering legal and administrative difficulties, how easy or difficult do you think it would be for the U.S. government to modify the tax system so that income taxes are based not only on income but also on personal characteristics?

[Very easy; Easy; Somewhat easy; Neither easy nor difficult; Somewhat difficult; Difficult; Very difficult]

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

[Never; Only some of the time; Most of the time; Just about always]

Screen Break

Some groups in society earn lower incomes than others, as you saw previously. For example, individuals with visual impairments tend to earn lower incomes, and so do individuals whose highest education level is a high school diploma or less.

In your opinion, what best explains why some groups in society earn lower incomes than others?

I believe that some groups earn lower incomes ...

[Mainly because they work less hard (not because they are less lucky); Because they work less hard and are less lucky; Mainly because they are less lucky (not because they work less hard)]

Screen Break

[**Block 6: Background Questions - 2**]

The study is almost over. We would like you to answer a few more questions about your background.

What is the 5-digit ZIP code of the place you currently reside in?

(Please enter a 5-digit numeric value)

What is your marital status?

[Single; Cohabitating couple; Married; Legally separated or divorced; Widowed]

How many dependents do you have?

Dependents are children or relatives that rely on you financially.

[None; 1; 2; 3 or more]

Are you a U.S. citizen?

[Yes; No]

Are you of Hispanic, Latino, or Spanish origin?

[Yes; No]

How would you describe your race?

[White; African American/Black; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; Mixed race; Other]

Screen Break

What is your current employment status?

[Full-time employee; Part-time employee; Self-employed or small business owner; Unemployed and looking for work; Unemployed and not looking for work; Student; Retiree]

[If "Full-time employee," "Part-time employee," or "Self-employed or small business owner"] Which of the following best describes the industry of your main occupation?

[If "Unemployed and looking for work," "Unemployed and not looking for work," or "Retiree"] Even if you are not currently working, which of the following best describes the industry of your latest occupation?

[Agriculture, Forestry, Fishing and Hunting; Mining; Utilities; Construction; Manufacturing; Wholesale and Retail Trade; Transportation and Warehousing; Information; Finance and In-

surance, and Real Estate, and Rental and Leasing; Professional, Scientific, and Management, and Administrative, and Waste Management Services; Educational Services, and Health Care and Social Assistance; Arts, Entertainment, and Recreation, and Accommodation and Food Services; Other Services, Except Public Administration; Public Administration; Military]

Screen Break

We now would like to ask you a few questions that relate to your health and the health of other members of your household (if any).

Do you or anyone in your household have serious difficulty seeing even when wearing glasses?

[Yes; No; Prefer not to answer]

Do you or anyone in your household have a physical or mental disability that limits their being employed, or substantially limits one or more of their major life activities, such as performing manual tasks, walking, speaking, breathing, learning, and working?

[Yes; No; Prefer not to answer]

Screen Break

What do you consider to be your political affiliation?

[Republican; Democrat; Independent; Other; Non-Affiliated]

Screen Break

End of survey

Thank you for participating in this study.

Please click the next button to finish the study. You will be automatically redirected.

References

Bartling, B. & Srinivasan, K. (2025), Paternalistic interventions: determinants of demand and supply, Technical report, CESifo Working Paper.

Boadway, R. & Pestieau, P. (2006), 'Tagging and redistributive taxation', *Annales d'Economie et de Statistique* pp. 123–147.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E. & Camerer, C. (2023), 'Econographics', *Journal of Political Economy Microeconomics* **1**(1), 115–161.

Lipovetsky, S. (2021), 'Game theory in regression modeling: A brief review on shapley value regression', *Model Assisted Statistics and Applications* **16**(2), 165–168.

Mankiw, N. G. & Weinzierl, M. (2010), 'The optimal taxation of height: A case study of utilitarian income redistribution', *American Economic Journal: Economic Policy* **2**(1), 155–176.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., Richards, S., Rodgers, R. & Schouweiler, M. (2024), 'Ipums usa: Version 15.0 [dataset]'.
URL: *https://doi.org/10.18128/D010.V15.0*

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

Francesco Capozza†
Krishna Srinivasan‡

Job Market Paper #2

November 5, 2025

[latest version] [Link to JMP #1]

Abstract

Evaluating the desirability of a reform typically involves weighing the gains of winners against the losses of losers using welfare weights, which measure the value that society places on a $1 increase in an individual's consumption. They can capture various normative ideals like utilitarianism and equality of opportunity. We elicit the welfare weights of the U.S. general population using experiments and show their robustness, validity, and temporal stability. We estimate an income elasticity of welfare weights between -0.78 and -0.70, which is roughly 5-9 times more progressive than the weights implied by U.S. tax and transfer policies. The optimal marginal tax rates based on our estimates of welfare weights are 28-30 percentage points higher, on average, than the current tax rates in the U.S.

Keywords: Welfare Weights, Policy Views, Income Taxation
JEL Classification: D63, H21, I31

*This work was supported by Grant 2104-31756 from the Russell Sage Foundation, by the University Research Priority Program "URPP Equality of Opportunity" of the University of Zurich, and by WZB. Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of the Russell Sage Foundation. We thank Sandro Ambühl, Aurelien Baillon, Björn Bartling, Katy Bergstrom, Alexander Cappelen, William Dodds, Ernst Fehr, Manuel Grieder, Nathaniel Hendren, Luca Henkel, Michael Kurschilgen, Michel Maréchal, Morten Støstad, Nina Serdarevic, Florian Scheuer, Itai Sher, Stefanie Stantcheva, Dmitry Taubinsky, Bertil Tungodden, and Roberto Weber for their helpful comments. We also thank participants at various conferences and seminars for their helpful comments. This study was pre-registered: https://doi.org/10.1257/rct.16001-1.1

†UB, IEB, CESifo. Email: francescocapozza94@gmail.com
‡UniDistance Suisse. Email: krishna.srinivasan@unidistance.ch.

1 Introduction

Most policy reforms create winners and losers. Evaluating the desirability of such reforms commonly involves weighing the gains of the winners against the losses of the losers using welfare weights. Welfare weights measure the value that society places on providing an additional dollar of consumption to any given individual. But which welfare weights should society use to evaluate reforms? Previous studies generally assume welfare weights (e.g., Saez 2001) or indirectly infer them from existing policies (e.g., Hendren 2020, Lockwood & Weinzierl 2016). In contrast, our paper takes a direct empirical approach by eliciting the welfare weights of the U.S. general population using experiments.

We develop a portable method to elicit welfare weights and apply this method using online experiments with samples of the U.S. general population ($N \approx 2000$). In the experiments, participants in the role of "Social Architects" face pairs of participants in the role of "Recipients." A Social Architect makes several real-stakes decisions that redistribute specified monetary amounts between the pairs of Recipients, with the real-world disposable incomes of the Recipients differing across pairs. The Recipients' disposable incomes span the U.S. income distribution. Social Architects' decisions reveal the welfare weights they implicitly assign to the Recipients.

Welfare weights are a reduced-form representation of underlying normative ideals, such as equality of opportunity, utilitarianism, distribution based on the source of income, and poverty alleviation (Saez & Stantcheva 2016). For example, a utilitarian Social Architect would assign welfare weights proportional to the Recipients' marginal utility of consumption. A Social Architect guided by equality of opportunity might assign higher welfare weights to Recipients from disadvantaged backgrounds. Our approach allows for evaluating policies conditioned on income using welfare weights—also conditioned on incomes—without the need to specify and uncover the underlying ideals.

Pooling responses of all Social Architects, we find that the median welfare weights are progressive: the median income elasticity of welfare weights is $\nu = -0.63$. This estimate implies that if a Social Architect assigns a welfare weight of 1 to a Recipient, they would assign a weight of 0.37 (about one-third) to a Recipient earning twice as much. In monetary terms, the Social Architect is indifferent between taking \$1 from a Recipient and giving \$0.37 to a Recipient earning half as much. Since these welfare weights are elicited based on Recipients' disposable incomes resulting from the current tax and transfer system, progressive welfare weights indicate a desire for redistribution beyond that achieved by the current system.

While the aggregate welfare weights are progressive, there is considerable heterogene-

ity. This heterogeneity is predicted by the Social Architects' background characteristics. Republicans have less progressive welfare weights relative to Democrats and Independents, consistent with the partisan gap in support for government redistribution documented in the literature (e.g., Kuziemko et al. 2015, Stantcheva 2021). We also find that Social Architects with above-median age have less progressive welfare weights.

We conduct several tests to assess the robustness, validity, and temporal stability of the elicited welfare weights. We assess the robustness of welfare weights to individual-level response quality using several proxies and estimate quality-adjusted measures of the progressivity, following Luttmer & Samwick (2018). We find that high-quality responses are more progressive; consequently, the quality-adjusted progressivity estimate is slightly more progressive than our unadjusted baseline estimate.

We also assess the robustness of welfare weights to variation in the features of the experimental design. First, we vary three features of the decision environment: (i) the visual placement of the two Recipients on the screen, (ii) the order in which Recipient pairs are presented, and (iii) the income of the Recipient common across the pairs of Recipients. Only the third feature affects welfare weights: weights are more progressive when the common Recipient earns \$60,000 rather than \$120,000. Second, we vary features of the elicitation method—a "staircase method" that presents participants with decisions in an adaptive manner. (iv) Randomizing the first decision reveals anchoring: a more progressive first decision leads to more progressive weights. (v) Randomizing whether Social Architects are prompted to consider the consequences of their choices affects welfare weights. Prompted Social Architects have slightly more progressive welfare weights, suggesting that some regressive choices are mistakes. We adjust our progressivity estimates for anchoring and mistakes and derive bounds from the treatments that vary the common Recipient's income.

The progressivity estimates obtained from our various robustness checks form a relatively tight bound: the implied income-elasticity of welfare weights lies in the interval $[-0.78, -0.70]$. Optimal policy formulas can implement these welfare weights using the parametric function c^{ν}, where c represents consumption and ν is a parameter governing the progressivity of the weights, with $\nu \in [-0.78, -0.70]$ or using constant relative risk aversion (CRRA) utilities with a coefficient of relative risk aversion given by $\gamma = [0.70, 0.78]$.

To validate the welfare weights, we test whether they correlate with two survey items measuring (i) general support for government redistribution and (ii) support for government redistribution at the margin. Social Architects with more progressive welfare weights express stronger support for government redistribution on both measures. When comparing levels, we find that 78% of Social Architects assign progressive welfare weights—implying a preference to redistribute at the margin—closely matching the 74% who support additional

progressive government redistribution at the margin. The close alignment between welfare weights and support for government redistribution, both in individual-level correlations and in aggregate shares, supports the validity of our measure of welfare weights. It also suggests that calibrating optimal policy formulas using the elicited welfare weights would likely result in policies with broad public acceptance.

As a second validation check, we ask whether the elicited welfare weights capture fairness concerns, as intended. We measure fairness concerns as perceived fairness of the current income distribution. We find that welfare weights capture fairness concerns, validating our measure of welfare weights. However, welfare weights also capture beliefs about taxes hurting the economy and views about the scope of the government. Reassuringly, welfare weights do not capture misperceptions about the level of taxes paid by individuals and the share earning low incomes (below $15,000).

To gauge the temporal stability of welfare weights, we re-elicited Social Architects' welfare weights in a second survey wave conducted four weeks after the first. The elicited welfare weights are temporarily stable, both at the individual and aggregate levels. The across-wave correlation in welfare weights is comparable to the correlations documented in the broader preference-elicitation literature and to the correlations of two survey measures of support for government redistribution that we fielded in both waves. Our finding of a high degree of temporal stability at the aggregate level supports using the elicited aggregate progressivity measures to calibrate optimal policy formulas.

We compare the welfare weights obtained from our experiment to the weights implied by the U.S. income tax schedule (obtained from Hendren (2020)) and tax and transfer policies (obtained from Hendren & Sprung-Keyser (2020)). These weights—referred to as "inverse optimum weights"—represent politicians' aggregation of citizens' welfare weights, possibly influenced by political economy considerations. The welfare weights obtained from our experiment are about 8 to 9 times more progressive than the inverse optimum weights implied by the tax schedule and about 5 times more progressive than those implied by tax and transfer policies.

Finally, we explore the implications of the elicited welfare weights for optimal non-linear income taxes in the U.S. We calibrate the optimal non-linear labor income tax formula provided by Saez (2001) using the estimates of welfare weights from our experiment. We find that the optimal marginal tax rates obtained based on our experimental estimates of welfare weights are 28-30 percentage points higher, on average, than the current tax rates in the U.S.

Our paper is related to four strands of literature. The first is the experimental literature that aims to identify the ideals that guide people's support for redistribution (e.g., Drenik &

Perez-Truglia 2018, Almås et al. 2020, Saez & Stantcheva 2016). With the exception of Saez & Stantcheva (2016), these papers do not elicit welfare weights. Saez & Stantcheva (2016) elicit welfare weights as a function of disposable income and taxes and use them to calibrate optimal linear income taxes with no behavioral responses. Our paper estimates weights as a function of disposable income alone, making them portable across policy domains, and uses them to calibrate optimal non-linear income taxes that incorporate behavioral responses. We also document robustness, validity, and temporal stability of welfare weight, supporting their use in optimal policy formulas.[1]

A related strand of literature aims to identify distributional preferences (e.g., Fisman et al. 2007, 2023, Fehr et al. 2024), and preferences for government redistribution (Alesina & Angeletos 2005, Luttmer & Singhal 2011, Cruces et al. 2013, Durante et al. 2014, Karadja et al. 2017, Kuziemko et al. 2015, Stantcheva 2021).[2]

The second strand of literature identifies the inverse-optimum welfare weights implied by the tax schedule (e.g., Hendren 2020, Lockwood & Weinzierl 2016, Jacobs et al. 2017, Bourguignon & Spadaro 2012) or by tax and transfer policies (Hendren & Sprung-Keyser 2020). In contrast, our paper adopts an experimental approach to directly elicit the welfare weights of citizens. An important limitation of inverse-optimum welfare weights is that they may not be normatively appealing if they are influenced by political economy considerations, such as lobbying (Stantcheva 2016, Lockwood & Weinzierl 2016).

The third strand of literature aims to incorporate normative ideals in optimal policy formulas by modifying individuals' utilities or the objective function (e.g., Weinzierl 2014, Fleurbaey & Maniquet 2006). In contrast, our paper elicits the welfare weights of the general population, which can capture various underlying ideals, and uses these weights to calibrate "standard" optimal policy formulas.

The paper is organized as follows. Section 2 outlines the theoretical framework. Section 3 describes the experimental design. Section 4 reports the experimental results. Section 5 compares the elicited welfare weights to those implied by tax and transfer policies. Section 6 examines their implications for optimal income taxes. Section 7 presents a discussion.

[1]There are also two methodological differences. First, Saez & Stantcheva (2016) rely on a non-representative sample recruited from Amazon Mechanical Turks, while our paper recruits samples broadly representative on age, gender, income, education, and region. Second, their elicitation uses hypothetical decisions, whereas ours involves real-stakes decisions. Our paper shows the importance of incentivizing choices.

[2]Papers such as Fisman et al. (2007) and Fisman et al. (2023) elicit the weights assigned by participants on themselves relative to a stranger using choices that involve a tradeoff between equity and efficiency. However, these weights conflate the role of self-interest motives, which are normatively unappealing in the welfare economics tradition (e.g., Cappelen et al. 2013). Moreover, these welfare weights are not elicited as a function of income (since the income of the stranger is not specified); consequently, they cannot be directly used to evaluate policies conditioned on incomes.

2 Theoretical Framework

We present a simple theoretical framework, adapted from Saez & Stantcheva (2016), to conceptualize welfare weights and describe the approach to elicit them through experiments.

2.1 Setup

Consider a population of N *Recipients* indexed by j. A Recipient earns income z_j, pays taxes $T(z_j)$ as a function of income, and consumes $c_j = z_j - T(z_j)$. A Recipient j's utility function is quasi-linear in consumption and given by $u_j = z_j - T(z_j) - v(z_j, \alpha_j)$, where v represents Recipients' disutility of work as a function of their income and various personal characteristics α_j (e.g., disability status). We consider the simple case where earnings are completely inelastic to taxes and transfers to focus on redistributive issues (the focus of our paper) rather than behavioral responses (absent in the experiment).

2.2 Welfare Weights

The *Social Architect* assigns *generalized social marginal welfare weights* (henceforth *welfare weights*) to the Recipients. The welfare weight $g_j = g(c_j, \theta_j)$ measures the Social Architect's assessment of the value of increasing the consumption of Recipient j by \$1.[3] These welfare weights are a function of the Recipients' consumption c_j and characteristics contained in θ_j (e.g., disability status or parental income). Welfare weights are defined up to a multiplicative constant, as they measure the relative value of consumption of Recipient j.

Welfare weights are a reduced-form representation of underlying normative ideals rather than being deep structural primitives. They can capture various ideals such as equality of opportunity, utilitarianism, redistribution based on the source of income, and poverty alleviation. For example, a utilitarian Social Architect would assign welfare weights proportional to the Recipients' marginal utility of consumption (captured by c_j). A Social Architect guided by equality of opportunity would assign higher welfare weights to Recipients from disadvantaged backgrounds (captured by θ_j).

Although welfare weights may vary with characteristics θ_j, we focus on the average welfare weights at each consumption level c, since reforms are conditioned solely on incomes. A Social Architect's average welfare weight, $\bar{g}_j$, at income level $c = c_j$ equals

$$\bar{g}_j = \frac{\sum_{j:c_j=c} g_j}{h(c)} \tag{1}$$

[3]In particular, a Social Architect maximizes a social welfare function (not necessarily welfarist), and the welfare weight g_j measures the marginal increase in the Social Architect's objective function that arises from increasing Recipient j's consumption by \$1.

where $h(c)$ is the number of Recipients with income $c = c_j$. A distribution of mean welfare weights can be consistent with multiple underlying ideals. For example, under certain assumptions, welfare weights decreasing with Recipients' incomes can be consistent with utilitarianism and equality of opportunity.

2.3 Identifying Welfare Weights

Consider a setting with two Recipients with consumption levels c_l and c_h, such that $c_h > c_l$. Consider a "small" (not necessarily budget-neutral) reform $(r_l, -r_h)$ where r_l is the amount given to the low-income Recipient and r_h is the amount taken from the high-income Recipient. We assume that the reform does not affect earnings (which are assumed to be inelastic to taxes) or tax liabilities: consequently, a \$1 transfer translates to a \$1 increase in consumption.[4]

If a Social Architect assigns average welfare weights $\bar{g}_l$ and $\bar{g}_h$ to the low-income and high-income Recipients, respectively, she values the reform at $\bar{g}_l \cdot r_l - \bar{g}_h \cdot r_h$. To identify the ratio of welfare weights $\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l}$, we identify the reform $(r_l, -r_h)$ that makes the Social Architect indifferent between implementing it and maintaining the status quo $(0, 0)$. Equating the Social Architect's valuation of this reform to her valuation of the status quo, we have $\bar{g}_l \cdot r_l - \bar{g}_h \cdot r_h = 0$. This implies,

$$\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l} = \frac{r_l}{r_h}. \tag{2}$$

The ratio of the welfare weights is inversely proportional to the ratio of the reform amounts. Equation (2) allows us to identify $\bar{g}_h$ and $\bar{g}_l$ up to a multiplicative constant. If $\tilde{g} < 1$, the welfare weight on the high-income Recipient is lower than the welfare weight on the low-income Recipient ($\bar{g}_h < \bar{g}_l$), which corresponds to *progressive* welfare weights. Analogously, $\tilde{g} > 1 (\bar{g}_h > \bar{g}_l)$ corresponds to *regressive* welfare weights and $\tilde{g} = 1$ ($\bar{g}_h = \bar{g}_l$) implies equal weights.

Equation (2) shows that a Social Architect's welfare weights can be elicited using their choices between various reform amounts by finding their preferred reform that leaves them indifferent between the reform and the status quo. This approach allows us to elicit welfare weights without uncovering the underlying normative ideals guiding the Social Architect. The next section presents the experimental design used to elicit welfare weights.

[4] Assuming that reforms do not affect tax liabilities (and that Social Architects believe this) helps identify welfare weights unconfounded by the effect of the reforms on government revenue.

3 Experimental Design and Sample

This section describes the sample and experimental design used to elicit the welfare weights of the U.S. general population and to test their robustness, validity, and temporal stability. The experimental design, sample restriction, and analyses were pre-registered.[5] Minor deviations from the pre-registration are discussed in Appendix Section B. The complete set of instructions can be found in Appendix Section G.

3.1 Eliciting Welfare Weights

3.1.1 Approach

Participants in our experiment are assigned to one of two roles: *Social Architect* or *Recipient*. Social Architects decide whether to implement various "reforms" that redistribute specified monetary amounts between pairs of Recipients, with the real-world disposable incomes of the Recipients differing across pairs. Their decisions identify the welfare weights they implicitly assign to Recipients. Social Architects' decisions may have real consequences: one randomly selected decision made by one randomly selected Social Architect is implemented. Social Architects also learn that the Recipients (i) will be randomly drawn from the U.S. general population, (ii) are at least 18 years old, and (iii) are U.S. citizens. Recipients are passive subjects who receive payments based on the Social Architects' choices.

Our approach has four key features. First, following the small reform approach to taxation (Saez 2001, Saez & Stantcheva 2016), the reforms in our experiment constitute a small one-time transfer. Second, to ensure that welfare weights are unconfounded by factors such as trust in government, the reforms in the experiment make no references to the government or real-world policies. Third, the reforms are presented in a gain-loss frame—one Recipient loses while the other gains—to mirror trade-offs in policy evaluation (Saez & Stantcheva 2016, Hendren & Sprung-Keyser 2020). Finally, to capture Social Architects' assessments of the value of transfers given the current tax and transfer system, we present them with Recipients' disposable incomes accrued from the current system.[6]

3.1.2 Decisions

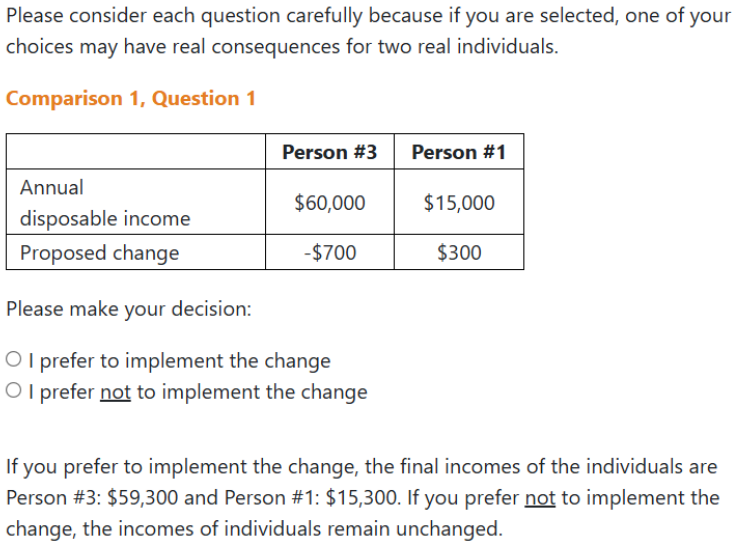
To elicit the welfare weights assigned by a Social Architect to a pair of Recipients, we present the Social Architect with several binary choices between implementing a reform and maintaining the status quo. Implementing a reform $(r_l, -r_h)$ would increase the income of the low-income Recipient by r_l and decrease the income of the high-income Recipient by r_h,

[5] See https://doi.org/10.1257/rct.16001-1.1.

[6] Eliciting welfare weights based on pre-tax incomes is challenging because it requires Social Architects to ignore the current disposable incomes of the Recipients, despite their incentivized decisions affecting these disposable incomes.

while maintaining the status quo leaves both incomes unchanged. To allow for negative transfers, we endow both Recipients with an initial endowment of $1,000, which is incorporated in their disposable incomes, and explicitly disclose this to the Social Architects. As discussed in the previous section, the reform $(r_l, -r_h)$ that makes a Social Architect indifferent between implementing it and maintaining the status quo, when plugged into Equation (2), reveals their welfare weights $\tilde{g} = \frac{\bar{g}_h}{\bar{g}_l} = \frac{r_l}{r_h}$.

Figure 1 presents a screenshot of one of the decisions presented to the Social Architects. We randomize the placement of the Recipients on the screen (left vs. right) across Social Architects. We refer to a reform as a "proposed change" instead of "reform" or "policy" to avoid any references to the government. At the bottom of the screen, we highlight how the reform affects the Recipients' incomes.


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

Comparison 1, Question 1

	Person #3	Person #1
Annual disposable income	$60,000	$15,000
Proposed change	-$700	$300

Please make your decision:

○ I prefer to implement the change
○ I prefer not to implement the change

If you prefer to implement the change, the final incomes of the individuals are Person #3: $59,300 and Person #1: $15,300. If you prefer not to implement the change, the incomes of individuals remain unchanged.

Figure 1: Screenshot of a Decision Presented to Social Architects

Table 1 presents the full set of reforms $(r_l, -r_h)$ used to elicit welfare weights and the corresponding $\tilde{g}$. To cover a wide range of $\tilde{g}$ without requiring extremely large transfers, we simultaneously vary r_l and r_h across the reforms.[7] The reform amounts range from $50 to $950 in increments of $50, with the average amount ($500) comparable to several tax and transfer policies in the U.S. (see Appendix Table A2). The reform in Row 10 corresponds to $\tilde{g} = 1$; the reforms above Row 10 correspond to $\tilde{g} < 1$, while those below Row 10 correspond to $\tilde{g} > 1$. The reforms above Row 10 are symmetric to those below Row 10, ensuring that the set of progressive weights is the inverse of the set of regressive weights: $\tilde{g}$ in Row $10 - n$ is equal to $\frac{1}{\tilde{g}}$ in Row $10 + n$, $\forall n \in 1, \ldots, 9$.

[7]For instance, to obtain $\tilde{g} = 19$ (the largest $\tilde{g}$ in the table) with a fixed $r_h = \$500$ requires $r_l = \$9{,}500$.

Table 1: Set of Reforms Used to Elicit Welfare Weights

Row	Reform $(r_l, -r_h)$	$\tilde{g}$
1	(\$50, −\$950)	0.05
⋮	⋮	⋮
6	(\$300, −\$700)	0.43
⋮	⋮	⋮
8	(\$400, −\$600)	0.67
9	(\$450, −\$550)	0.82
10	(\$500, −\$500)	1
11	(\$550, −\$450)	1.22
12	(\$600, −\$400)	1.5
⋮	⋮	⋮
14	(\$700, −\$300)	2.33
⋮	⋮	⋮
18	(\$900, −\$100)	9
19	(\$950, −\$50)	19

Notes: The table presents the set of reforms used to elicit welfare weights. A reform $(r_l, -r_h)$ increases the income of the low-income Recipient by r_l and decreases the income of the high-income Recipient by r_h. $\tilde{g} = \frac{r_l}{r_h}$ is the ratio of the reform amounts.

We aim to identify a Social Architect's switch-point—the row where they switch from preferring the status quo (where no reform is implemented) to preferring the reform. The mid-point of the reforms in this row and the previous row represents the reform at which a Social Architect is just indifferent between implementing it and maintaining the status-quo.[8] To identify a Social Architect's switch-point, we use a "staircase method," which presents them with three to five decisions that are adaptively selected based on their previous choices. This method allows us to obtain precise weights with few decisions per participant.[9] We

[8]If a Social Architect always prefers a reform (i.e., switches in the first row), we take the mid-point of (\$50, −\$950) and (\$0, −\$1000), with the latter (corresponding to $\tilde{g} = 0$) giving the minimum possible amount to the low-income Recipient. If a Social Architect always prefers the status quo (i.e., never switches), we take the mid-point of (\$950, −\$50) and (\$1000, −\$0), with the latter reform (corresponding to $\tilde{g} = \infty$) taking the minimum possible amount from the high-income Recipient.

[9]Compared to the Multiple Price List (MPL) method, which presents all decisions simultaneously on a single page, the staircase method offers several advantages. First, the staircase method is simpler for the general population to understand; its simplicity is evidenced by Falk et al. (2018), who implemented the method to elicit risk and time preferences in nationally representative samples across the world. Second, the staircase method can pin down preferences with fewer decisions, reducing the burden on Social Architects. Third, it avoids the bias associated with the order of presentation of the decisions on the screen—a bias

randomize the first decision in the staircase to assess the sensitivity of the results to the starting point, following the approach of Luttmer & Samwick (2018) and Bursztyn et al. (2023). For half the Social Architects, the reform in the first decision is ($300, −$700), while for the other half, it is ($700, −$300). We discuss this and other randomizations in more detail in the next section. Appendix Figures A2 and A3 present a graphical representation of the staircase method with the two different starting points.

3.1.3 Recipients' Incomes and Comparisons

Social Architects are presented with six Recipients whose disposable incomes are chosen to span the U.S. disposable income distribution while satisfying three design constraints: (i) the highest income is capped at $500,000 to keep the recruitment of high-earners feasible, (ii) the ratios of neighboring incomes are constant to minimize cognitive burden, and (iii) incomes are rounded to the nearest $10,000 to reduce errors due to numeracy (e.g., List et al. 2023, Strulov-Shlain 2023). The resulting incomes range from $15,000 to $480,000, with each subsequent income double the previous one (see Appendix Figure A4 for a graphical representation).

A Social Architect is presented with five comparisons, with the incomes of the pairs of Recipients varying across the comparisons (see Table 2). To check the robustness of the welfare weights to the order in which comparisons are presented, we randomize the order—half the Social Architects see income differences in ascending order (as per the table), while the other half see them in descending order. A common Recipient appearing across all comparisons allows us to identify the relative welfare weights assigned to the six Recipients. To check the robustness to the common Recipient, we randomize the income of the common Recipient to be either $60,000 (as per the table) or $120,000.

Table 2: Recipients' Incomes Across Comparisons

	Comparison				
Recipient	1	2	3	4	5
Low-Income	$15,000	$30,000	$60,000	$60,000	$60,000
High-Income	$60,000	$60,000	$120,000	$240,000	$480,000

3.1.4 Incentives

To incentivize thoughtful decisions, Social Architects are informed that one participant in the study would be randomly selected at the end of the study, and one of their decisions would be randomly selected and implemented. The decision could be selected from either

observed in the MPL method (e.g., Jack et al. 2022). The limitations of the staircase method are discussed in Section 3.2.2.

of the two waves of data collection (discussed in the next section). Each decision saliently reminds Social Architects of the incentives (see Figure 1). Results from an additional study confirm the importance of incentivizing Social Architects' decisions.[10]

3.2 Assessing Robustness, Validity, and Temporal Stability of Welfare Weights

To assess the robustness, validity, and temporal stability of the elicited welfare weights, we conduct various tests.

3.2.1 Assessing Robustness to Quality of Responses

We assess the robustness of welfare weights to individual-level response quality using proxies of response quality. First, we flag Social Architects failing one or more comprehension questions on the first attempt as having low-quality responses. We included three comprehension questions that test whether Social Architects have understood (i) that they will make decisions regarding six Recipients, (ii) that they will be presented with disposable incomes (rather than pre-tax incomes), and (iii) that their decisions may have real stakes. Second, we elicit Social Architects' confidence in their decisions and classify those with low confidence as having low response quality. Third, we check decision consistency by including a sixth comparison identical to the third comparison (Recipients earning \$60,000 and \$120,000) and flag those whose welfare weights are inconsistent across comparisons as having low-quality responses. Finally, we flag completion times beyond two standard deviations from the mean within each treatment as an indicator of low response quality. Similar proxies have been used in the literature to assess response quality (e.g., Luttmer & Samwick 2018, Enke et al. 2023, Stantcheva 2023).

3.2.2 Assessing Robustness to Design Features

We assess whether the estimated welfare weights are robust to features of the experimental design by introducing several treatments that vary these features across Social Architects.

The first three treatment dimensions vary features of the decision environment, including (i) the placement of the Recipients on the screen (left vs. right), (ii) the order of presentation of the comparisons (ascending vs. descending order of income differences), and (iii) the Recipient common across the comparisons (\$60,000 vs. \$120,000). The set of high-to-low income ratios of Recipients is identical across treatments that vary the common Recipient, which allows us to isolate the role of the common Recipient.

The remaining two treatments vary features of the staircase method, addressing two potential limitations of the method. We randomize the reform in the first decision in the

[10] In a study conducted on Prolific ($N = 1965$), we find that Social Architects' welfare weights are more progressive with hypothetical decisions compared to real stakes decisions, consistent with an explanation of Social Architects considering the trade-offs more carefully in the latter. Appendix Section F presents the details of the study and analysis.

staircase to be either (\$300, −\$700) or (\$700, −\$300), following Luttmer & Samwick (2018) and Bursztyn et al. (2023). This allows us to test whether Social Architects anchor to the first decision. Finally, we vary whether Social Architects are prompted to consider the consequences of their choices to test whether their choices are based on mistakes. Prompted Social Architects are asked to choose between two final income distributions, which reflect the consequences of the two options they were presented with. They can only proceed if their preference on whether to implement the reform aligns with their preferred final income distribution.[11] To minimize the burden on Social Architects, they are only prompted in the first decision of each comparison because a mistake in the first decision leads to the largest distortion in the estimated welfare weights.[12] Prompted Social Architects are informed that they will be prompted in the first decision of each comparison. Appendix Figure A1 presents a screenshot of the prompt.

3.2.3 Correlation with Support for Redistribution

We validate our measure of welfare weights by testing whether they correlate with measures of support for government redistribution. We measure general support for government redistribution with a question used in the General Social Survey (GSS) and by Alesina et al. (2018) and support for government redistribution at the margin (i.e., beyond the current tax and transfer system). The order of the questions is randomized across Social Architects.

3.2.4 Do Welfare Weights Capture Fairness Concerns?

We test whether welfare weights capture fairness concerns (as intended) or other concerns using additional survey questions, drawn from Stantcheva (2021), capturing (i) fairness of the current income distribution, (ii) trickle-down economics, (iv) beliefs about labor distortions from taxes, (v) trust in government, and (vi) views about the scope of government. The order of the questions is randomized across Social Architects. Motivated by recent work documenting misperceptions (e.g., Hvidberg et al. 2023, Rees-Jones & Taubinsky 2020, Kuziemko et al. 2015), we also test whether welfare weights capture misperceptions about the level of taxes paid by individuals and the share of low-income individuals. Detailed descriptions of the measures are presented in Appendix Section A.

[11] This approach builds on Bursztyn et al. (2023), who ask participants for their agreement with a statement that reformulates their choices and provides them with the opportunity to revise their choices, and papers that nudge participants to reconsider their choices (e.g., Burchardi et al. 2021, Berry et al. 2020, Abdellaoui et al. 2019).

[12] As the staircase converges toward a Social Architect's preference, mistakes made later in the sequence lead to smaller distortions in the true welfare weights. When a mistake occurs, the true ratio of welfare weights $\tilde{g}$ is scaled by a factor of x and the observed ratio is $x\tilde{g}$. If a mistake occurs in the first decisions (with no subsequent mistakes), the variance of x is 17.98. The variance drops to 6.20, 1.53, 0.59, and 0.51 when the mistake occurs in the second, third, fourth, or fifth decision, respectively. Thus, mistakes in earlier decisions lead to larger deviations between observed and true welfare weights.

3.2.5 Temporal Stability of Welfare Weights

To gauge the temporal stability of welfare weights, we re-elicited Social Architects' welfare weights in a second survey wave conducted four weeks after the first. Participants with complete responses in Wave 1 were invited to Wave 2. The welfare weights elicitation protocol (including treatment assignments) was identical across waves. The short interval limits the scope for external shocks or changes in personal circumstances to alter preferences (Chuang & Schechter 2015). We benchmark the stability of welfare weights against the stability of our two measures of support for government redistribution, which were included in both waves.

3.3 Data Collection

We conducted the study across two waves. In Wave 1 of data collection, we recruited participants in the role of Social Architects from the data collection provider Prolific.[13] Recruitment was based on three demographic quotas available on Prolific: gender, political affiliation, and age. We implemented the experiment using oTree (Chen et al. 2016). The data collection for Wave 1 began on 12 May 2025 and lasted three days. Our final sample includes 1,996 participants.[14] The median completion time is 15.7 minutes.

We conducted the second wave of data collection roughly four weeks after the completion of the first wave. We invited 1986 participants to Wave 2, including those who completed the survey in the first wave and had a valid ID prior to data collection. The data collection for Wave 2 began on 9 June 2025 and lasted three days. With a follow-up rate of 70%, our final sample includes 1397 participants.[15] The median completion time is 9 minutes.

3.4 Summary Statistics

Table 3 presents the sample averages of Social Architects' background characteristics and the population averages. The Wave 1 sample broadly resembles the population but under-represents very high-income earners ($\geq$ \$150,000) and individuals with a high-school education or less, while over-representing very low-income earners ($<$ \$30,000) and those with a Bachelor's degree. We use sampling weights that help align sample means with population means; after weighting, the absolute differences in the sample and population means for all background characteristics are less than 0.1 percentage points. The Wave 1 and Wave 2 samples are also very similar.[16]

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

[14] We recruited 2,194 participants in Wave 1. Following our pre-registered sample selection, we drop participants with multiple survey responses (0.18%), who failed the attention check (0.42%), and who did not complete the study.

[15] We invited 1,986 participants in Wave 2. Out of these, 1,447 participants took our survey (response of 73%). Following our pre-registered sample selection, we drop participants with multiple survey responses (0.35%), who failed the attention check (0.14%), and who did not complete the survey.

[16] Many characteristics are not balanced across treatment arms within each wave. Appendix Table A1 reports these balance tests. We control for background characteristics in the regressions exploring treatment

Table 3: Summary Statistics of Sample and Population

Variable	Wave 1	Wave 2	Population
Male	0.49	0.49	0.49
Income $<$ 30,000	0.26	0.28	0.13
Income 30–59,999	0.29	0.28	0.18
Income 60–99,999	0.26	0.26	0.23
Income 100–149,999	0.12	0.12	0.20
Income $\geq$ 150,000	0.07	0.07	0.26
Education: High school or less	0.11	0.11	0.37
Education: Some college	0.18	0.19	0.20
Education: Bachelor or Associate	0.45	0.47	0.30
Education: Masters or above	0.26	0.24	0.13
Age: 18–24	0.12	0.09	0.11
Age: 25–34	0.18	0.16	0.17
Age: 35–44	0.17	0.17	0.17
Age: 45–54	0.15	0.16	0.16
Age: 55–64	0.25	0.27	0.16
Age: $\geq$ 65	0.13	0.14	0.23
Region: Northeast	0.16	0.17	0.17
Region: Midwest	0.18	0.18	0.20
Region: South	0.44	0.43	0.39
Region: West	0.22	0.22	0.24
Republican	0.32	0.30	0.28

Notes: The table presents the average background characteristics of our sample and the U.S. population. Columns (2) and (3) report the sample demographics based on 1996 participants in Wave 1 and 1397 participants in Wave 2, respectively. Column (4) reports U.S. population demographics, computed using the 2023 American Community Survey (ACS) 1-year estimates for individuals aged 18 and older. The population share of Republicans is the average share of individuals identifying as Republican, based on the Gallup poll in 2024.

We also test representativeness using two survey items fielded in our experiment with wording identical to that used in large national polls. First, we find that 20% of the weighted Wave 1 sample report that the government can be trusted "most of the time" or "just about always," closely matching the share (22%) in the Pew Research Center's 2024 national poll. Second, 56% of the weighted Wave 1 sample supports reducing income differences between the rich and the poor, which is very similar to the corresponding share (53%) in the 2024 General Social Survey (GSS).[17] These close parallels to large-scale national surveys,

effects.

[17]Responses range on a scale from 1 to 7, with higher values indicating stronger preferences for reducing income differences between the rich and poor. We code values above 4 as supporting reductions in income differences. Unlike the GSS question, our question does not include a "Don't know" option. Appendix

particularly along preferences for redistribution, provide additional reassurance regarding the representativeness of our sample.

4 Welfare Weights of the General Population

4.1 Data Description

We begin by exploring the distribution of $\tilde{g}$—the ratio of welfare weights assigned to the high-income and low-income Recipients—across the six pairs of Recipients. We infer $\tilde{g}$ from each Social Architect's switching behavior using Equation (2). Figure 2a presents the distribution of $\tilde{g}$ using data from Wave 1 in treatments where the common Recipient across comparisons earns \$60,0000, while Figure 2b displays the analogous distribution when the common Recipient earns \$120,000.

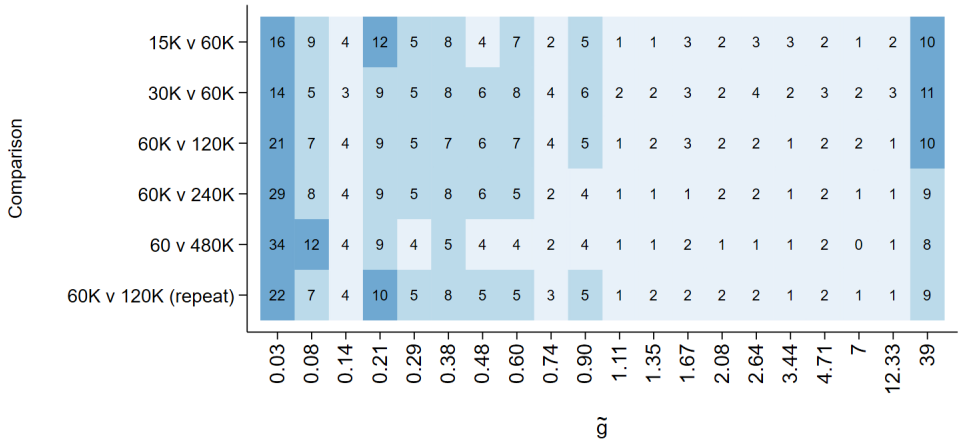
There are several patterns in the data. First, there is substantial heterogeneity in $\tilde{g}$: within each comparison, $\tilde{g}$ ranges from 0.03 (most progressive) to 39 (most regressive). Second, despite the heterogeneity, the distribution of $\tilde{g}$ skews toward low values of $\tilde{g}$ in every comparison—skews toward progressivity. Third, this progressive skew is stronger in comparisons featuring larger income differences between Recipients, implying that the aggregate welfare weights are progressive. Fourth, there is bunching at the extremes: 14–34% of Social Architects bunch at the progressive extreme ($\tilde{g} = 0.03$), while 8–15% bunch at the regressive extreme ($\tilde{g} = 39$); relatively few bunch at the "egalitarian" values $\tilde{g} = 0.09$ or $\tilde{g} = 1.11$. Finally, the distribution of $\tilde{g}$ is similar across identical comparisons: in the comparison that appears in both treatments (incomes \$60,000 and \$120,000), a Kolmogorov–Smirnov test fails to reject the equality of the two distributions ($D = 0.024$, $p = 0.929$). We also see that the distributions for the third and sixth comparisons (with Recipients earning \$60,000 and \$120,000) are indistinguishable in treatments where the common Recipient earns \$60,000 (Kolmogorov–Smirnov test; $D = 0.021$, $p = 0.979$) and \$120 000 ($D = 0.016$, $p = 1.000$).

4.2 Progressivity of Welfare Weights

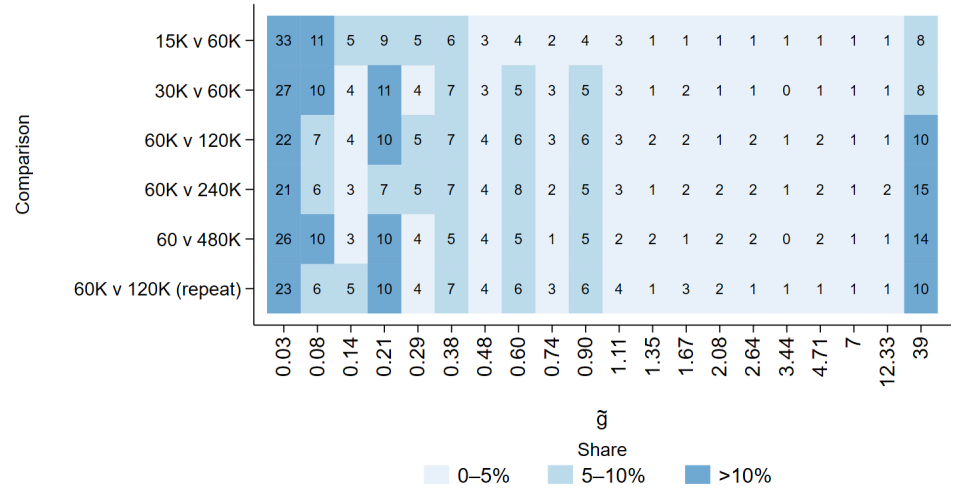
We find that 40% of Social Architects assign welfare weights that are weakly decreasing in Recipients' incomes (i.e., weakly progressive), while 8% assign weights that are weakly increasing (i.e., weakly regressive).[18] We estimate the progressivity of the Social Architects' welfare weights using a commonly employed parametric function (e.g., Saez 2002, Allcott et al. 2019): welfare weights g_r assigned by a Social Architect to Recipients (indexed by r) with disposable incomes c_r is proportional to c_r^{ν}, where ν is a parameter governing the

Figure A6 presents the distribution of responses of our sample.

[18]Non-monotonic weights are not necessarily due to measurement error. They could also be due to various normative ideals. For example, some Social Architects may be broadly progressive but assign a lower weight to very low-income individuals, whom they perceive as "lazy" (Drenik & Perez-Truglia 2018).



(a) Common Recipient earns \$60K



(b) Common Recipient earns \$120K

Figure 2: Distribution of $\tilde{g}$ across Comparisons and Treatments

Notes: The figures present the distribution of $\tilde{g}$—the ratio of welfare weights assigned to the high-income and low-income Recipients—across the six comparisons with the six pairs of Recipients. Each cell presents the share of participants with a given $\tilde{g}$ in a given comparison. The shares sum to 100% in each row. The figures in both panels use data from Wave 1; Panel (a) uses data from treatments in which the common Recipient across pairs earns \$60,000, while Panel (b) uses data from treatments in which the common Recipient earns \$120,000.

progressivity of the welfare weights. A negative value ($\nu < 0$) indicates progressive weights, while a positive value ($\nu > 0$) implies regressive weights. With a utility function that is additively separable in consumption and leisure and exhibits constant relative risk aversion in consumption, $-\nu$ corresponds to the coefficient of relative risk aversion.

To estimate ν for each Social Architect, we model the conditional expectation of the assigned welfare weights as

$$E[g_r \mid c_r] = exp(\alpha + \nu log(c_r)), \tag{3}$$

where the parameter α is a constant. The progressivity parameter ν can be interpreted as the income elasticity of welfare weights. We estimate ν using a Poisson Pseudo Maximum Likelihood (PPML) estimation.[19] We aggregate the welfare weights across Social Architects by computing the median estimate of ν—the median preference corresponds to the aggregate preference under the median voter theorem (Downs 1957).

Pooling the data across all treatments in Wave 1 and weighting the sample using sampling weights to match the sample with the population, we find a median progressivity estimate of $\nu = -0.63$ (see Figure 3).[20] In practical terms, this elasticity implies that if a Social Architect assigns a welfare weight of 1 to a Recipient, they would assign a weight of 0.37 (about one-third) to a Recipient earning twice as much. In monetary terms, this suggests that the Social Architect is indifferent between taking \$1 from the higher-income Recipient (decreasing welfare by $0.37 \cdot \$1$) and giving \$0.37 to the lower-income Recipient (increasing welfare by $1 \cdot \$0.37$). Since these welfare weights were elicited given the current tax and transfer system, our finding of progressive welfare weights implies that the Social Architects desire additional redistribution at the margin.

While the aggregate welfare weights are progressive, there is considerable heterogeneity across Social Architects. Specifically, 78% of Social Architects assign progressive weights ($\nu < 0$), while 22% assign regressive weights ($\nu > 0$). Applying sampling weights, these proportions remain similar: 75% progressive and 25% regressive. Additionally, there is significant variation in welfare weights within both of these groups (see Appendix Figure A5 for the distribution of ν). The minimum observed value of ν in our data is -4.18, and the maximum is 4.

4.3 Welfare Weights and Background Characteristics

We explore whether the observed heterogeneity in welfare weights is predicted by Social Architects' background characteristics. Table 4 presents median regression estimates of progressivity (ν) on background characteristics.

Our results indicate that Republicans assign less progressive welfare weights relative to

[19] We employ PPML estimation rather than OLS estimation (by log-linearizing Equation (3)), as the latter can be severely biased in the presence of heteroskedasticity (Silva & Tenreyro 2006).

[20] The unweighted median is slightly more progressive: $\nu = -0.71$, bootstrap SE $= 0.02$, 95% normal CI $= [-0.75, -0.66]$.

Democrats and Independents: the median ν among Republicans is 0.12 points higher than among non-Republicans. This result is consistent with the partisan gap in support for government redistribution observed in the literature (e.g., Alesina et al. 2018, Stantcheva 2021, Kuziemko et al. 2015). While Republicans assign less progressive weights, their weights are progressive, indicating that Republicans also desire additional redistribution at the margin.

Table 4: Welfare Weights and Background Characteristics

	(1)
Republican	0.118**
	(0.051)
High Income	0.076
	(0.049)
Male	0.023
	(0.047)
High Age	0.138***
	(0.048)
High Education	-0.093*
	(0.055)
Constant	-0.791***
	(0.044)
Observations	1998

Notes: The table presents coefficient estimates from a median regression. The dependent variable is the progressivity of the welfare weights (ν). *Republican* equals 1 for Republicans and 0 otherwise. *High Income* equals 1 for Social Architects with above median income and 0 otherwise. *High Age* equals 1 for Social Architects with above median age and 0 otherwise. *High Education* equals 1 for Social Architects with above median education and 0 otherwise. *Male* equals 1 if a Social Architect is male and 0 otherwise. The regression uses data from Wave 1. Standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Social Architects with above median income (above $54,000) assign less progressive weights, consistent with the income gap in support for redistribution observed in the literature (e.g., Singhal 2008, Cohn et al. 2023). However, the observed effect of income is not statistically significant. Although income is a weak predictor of progressivity, it predicts the weight assigned to the Recipient with similar incomes. Social Architects generally assign a higher weight to Recipients with incomes similar to their own relative to other Recipients, with the effects being largest for lower-income Social Architects (see Appendix Table A3).[21]

[21]The pattern is consistent with an explanation of self-interest motives. However, we cannot rule out alternative explanations, such as Social Architects having different preferences regarding Recipients with similar incomes.

Above-median aged (55 years or older) Social Architects assign less progressive welfare weights. We do not find a significant difference across education groups or across genders.

4.4 Assessing Robustness to Quality of Responses

We assess the robustness of the estimated welfare weights to individual-level response quality using several indicators of response quality. We flag potentially low-quality responses when participants (i) fail one or more comprehension questions on the first attempt, (ii) exhibit extreme completion times, (iii) provide internally inconsistent welfare weights (type clarification—i.e., progressive or regressive—differs across the identical third and sixth comparisons), or (iv) report low confidence in their decisions. About half the Social Architects are flagged by one or more indicators, with most of these Social Architects being flagged by only one indicator.[22] Reassuringly, Social Architects' welfare weights exhibit a high degree of internal consistency: about 90% of Social Architects have the same type classification (progressive or regressive weights) across identical comparisons. Furthermore, the distributions of $\tilde{g}$ are similar across the identical comparison (see Section 4.1).

Table 5 presents median regression estimates of progressivity (ν) on indicators of response quality. The explanatory variables also include treatment indicators and demographic controls. The median value of ν is 0.14 points higher among those who fail at least one comprehension question on their first try and 0.50 points higher among those with inconsistent welfare weights; Social Architects with low-quality responses tend to assign less progressive weights. We do not find a statistically significant effect of the other indicators.

We adjust the progressivity estimates for response quality following Luttmer & Samwick (2018). Each coefficient in Table 5 represents the median bias introduced by having a low-quality response. For each Social Architect whose answer is flagged as low-quality by a given indicator (indicator takes a value of 1), we subtract that indicator's estimated bias (the coefficient estimate) from the estimated ν. This correction yields unbiased estimates under two conditions: (1) differences in true underlying progressivity between those with high- and low-quality responses are fully captured by demographic controls (rather than by the quality indicators), and (2) the median ν among those with high-quality responses is measured without error.[23] Our adjustment uses the two proxies that have a statistically significant effect on progressivity. Applying this adjustment yields a weighted median progressivity estimate of $\nu = -0.71$, which is 0.08 points lower than our unadjusted baseline estimate (see

[22] About 37% of Social Architects are flagged by only one response-quality indicator, 11% by two indicators, 2% by three indicators, and 0.2% by all four indicators. Overlap across pairs of indicators is also limited (see Appendix Figure A8).

[23] Theoretical work shows that additive errors with a median of zero leave the population median unbiased (Schennach 2022, Hausman et al. 2021). Empirically, several studies report that the median error of estimates of interest is close to zero (Bollinger 1998, Bound & Krueger 1991).

Table 5: Welfare Weights and Response Quality

	(1)
Fail comprehension	0.145**
	(0.066)
Inconsistent	0.501***
	(0.067)
Extreme time	-0.057
	(0.096)
Low confidence	0.074*
	(0.043)
Observations	1996
Controls?	Yes

Notes: The table presents coefficient estimates from a median regression. The dependent variable is the progressivity of the welfare weights (ν). *Fail comprehension* equals 1 if a Social Architect failed one or more comprehension questions in the first try and 0 otherwise. *Inconsistent* equals 1 if, in the identical third and sixth comparisons, a Social Architect assigns progressive welfare weights ($\tilde{g} < 1$) in one and regressive welfare weights ($\tilde{g} > 1$) in the other. *Extreme time* equals 1 if a Social Architect's time spent on the survey lies beyond two standard deviations of the mean. *Low confidence* equals 1 if a Social Architect reports confidence levels lower than the highest category of "Very Much." Controls include treatment indicators (those specified in Table 6) and demographic controls (those specified in Table 4). The regression uses data from Wave 1. Standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Figure 3). The response-quality adjusted estimates are more progressive since low-quality responses exhibit less progressive welfare weights.

4.5 Assessing Robustness to Design Features

To assess the robustness of the welfare weights to the features of the experimental design, we implemented several treatments that vary the features across Social Architects. Table 6 presents median regression estimates of progressivity (ν) on treatment indicators. Column (1) includes only treatment indicators as explanatory variables, while Column (2) additionally includes background characteristics as controls.

The first three treatment dimensions vary the features of the decision environment. The placement of the Recipients on the screen (left vs. right) does not have a significant effect on the progressivity of weights. The order of the comparisons (ascending vs. descending order of income differences) has a small but statistically insignificant effect on the progressivity of the weights. In contrast, the common Recipient across comparisons matters: welfare weights are more progressive (ν is 0.13 to 0.15 points lower) when the common Recipient earns $60,000

rather than \$120,000, even though the set of high-to-low income ratios is identical across the two treatments. A plausible mechanism is the common Recipient's relative position: the \$60,000 common Recipient is the lower income Recipient in three comparisons, while the \$120,000 common Recipient is the lower income Recipient in only two comparisons.[24]

The remaining two treatment dimensions vary features of the staircase method, addressing two potential limitations of the method. We find that welfare weights are more progressive (ν is 0.18 points lower) when the first decision in the staircase is more progressive (\$300, −\$700) than when it is less progressive (\$700, −\$300), consistent with Social Architects anchoring to the first decision. We also find that welfare weights are more progressive (ν is 0.07 to 0.09 points lower) when Social Architects are prompted to consider the consequences of their choices compared to when they are unprompted, indicating that some regressive choices in the unprompted treatments are likely mistakes.[25] This effect is only marginally significant without controls but becomes significant at the 5% level once demographic controls are included.

We find very similar results when considering univariate regression in which the treatment indicators enter the regressions separately (see Appendix Table A4).

Reconciling Estimates Across Frames

Guided by the framework of Bernheim & Rangel (2009) and subsequent empirical studies, we apply several strategies to reconcile progressivity estimates across "frames" (treatments). The resulting alternative progressivity estimates are presented in Figure 3.

When frames differ in whether Social Architects are prompted to consider the consequences of their choice, we treat the unprompted frame as one in which some choices reflect mistakes and the prompted frame—because the prompt plausibly reduces such mistakes—as the "welfare-relevant" frame. This follows the common practice of recovering choice-mappings from frames in which mistakes are corrected (e.g., Allcott & Taubinsky 2015, Chetty et al. 2009, Ambuehl et al. 2022). The weighted median value of ν in the prompted frame is -0.75, about 0.12 points lower than our baseline estimate ($\nu = -0.63$).

For frames that vary the first decision of the staircase, we interpret anchoring to the first decision as a bias and correct the progressivity estimates for this bias (see Allcott et al. (2022)

[24]If people are more sensitive to disadvantageous than to advantageous inequality (Fehr & Schmidt 1999), the greater share of comparisons in which the common Recipient is worse off in treatments with the \$60,000 common Recipient relative to those with the \$120,000 common Recipient can tilt choices toward more progressive weights in the former.

[25]An alternative explanation is that the prompt induces experimental demand. To mitigate demand effects, we explicitly informed Social Architects before they began that they would be prompted in the first decision of each comparison. Moreover, we did not explicitly ask them to change their choice when there was an inconsistency (which might induce demand effects). Instead, we asked them to make their choice and indicate their preferred final distribution once again.

Table 6: Treatment Effects

	(1)	(2)
Low-income Recipient on Left	-0.044	-0.026
	(0.045)	(0.047)
Common recipient 60K	-0.131***	-0.150***
	(0.045)	(0.047)
Descending order of income diff	-0.076*	-0.045
	(0.045)	(0.047)
First decision (300, -700)	-0.188***	-0.182***
	(0.045)	(0.047)
Prompted	-0.075*	-0.093**
	(0.045)	(0.047)
Constant	-0.448***	-0.545***
	(0.055)	(0.071)
Observations	1996	1996
Controls?	No	Yes

Notes: The table presents coefficient estimates from a median regression. The dependent variable is the progressivity of the welfare weights (ν). The explanatory variables include treatment indicators and demographic controls (those specified in Table 4). The regressions use data from Wave 1. Standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

and Chetty et al. (2009) for a similar approach of inferring bias parameters). Following the adjustment strategy of Luttmer & Samwick (2018) and Bursztyn et al. (2023), we assume that the true values of the reform amounts that make Social Architects indifferent between the reform and the status quo are a weighted average of reported reform amounts and the reforms amounts in the first decision (details in Appendix Section E.1). We find that the extent of anchoring in each comparison is small; consequently, the median progressivity estimate after correcting for anchoring is $\nu = -0.66$, which is very similar to the baseline estimate.

Finally, when the frames vary the income of the common Recipient across comparisons, there is no obvious "welfare-relevant" frame. This is supported by results showing that indicators of response quality and temporal stability are not predicted by the frame (see Appendix Table A6). We embrace the normative ambiguity about the welfare-relevant frame and obtain bounds on the progressivity estimates (see Goldin & Reck (2022) for a similar approach). We estimate the weighted median value of ν in each frame that varies the common Recipient, and additionally, focus on the subset of prompted treatments to correct for mistakes. The weighted median value of ν when the common Recipient earns \$60,000 is

-0.78 and when the common Recipient earns \$120,000 is -0.70.[26]

Overall, the bounds on progressivity estimates, determined by the frames that vary the common Recipient and frames that prompt Social Architects to consider the consequences of their mistakes, are given by $\nu \in [-0.78, -0.70]$.

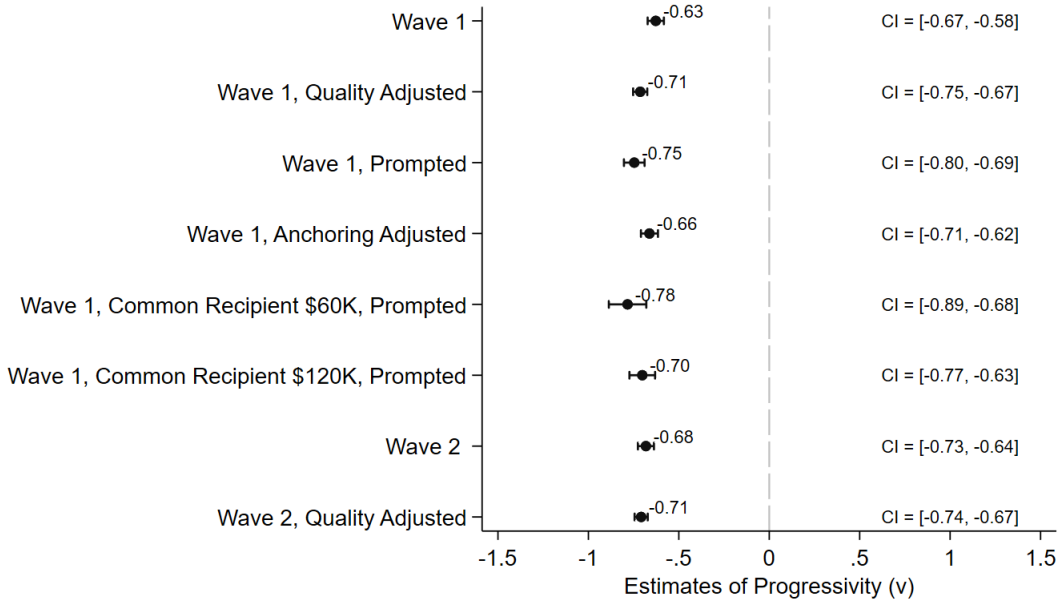


Figure 3: Estimates of Progressivity (ν)

Notes: The figure presents progressivity estimates ν from various specifications. All estimates are weighted using sampling weights to match the averages of the specified sample to the population averages. Confidence intervals are 95% normal confidence based on bootstrap standard errors.

4.6 Welfare Weights and Support for Redistribution

We validate our measure of welfare weights by testing whether it correlates with two measures of support for government redistribution. Social Architects with more progressive welfare weights should, all else equal, exhibit stronger support for government redistribution. We test these correlations using linear regressions presented in Table 7. In Column (1), the outcome variable is the Social Architects' support for reducing income differences between the rich and the poor (ranging from 1 to 7), with higher values indicating stronger support for redistribution. In Column (3), the outcome variable is support for redistribution beyond that achieved by the current tax and transfer system (ranging from -2 to 2), with positive (negative) values indicating redistribution from high-income (low/middle-income)

[26]Pooling over the prompted and unprompted treatments, the weighted median value of ν when the common Recipient earns \$60,000 is $\nu = -0.73$ (bootstrap SE $= 0.04$, 95% normal CI $= [-0.81, -0.65]$), and when the common Recipient earns \$120,000 is $\nu = -0.55$ (bootstrap SE $= 0.03$, 95% normal CI $= [-0.60, -0.49]$).

individuals to low/middle-income (high-income) individuals. The explanatory variable in both columns is the progressivity of the weights (ν). On average, Social Architects support increasing government redistribution on both measures (See Appendix Figures A6 and A7).

Social Architects with more progressive welfare weights (lower values of ν) exhibit stronger support for redistribution on both measures, and the associated coefficients are highly statistically significant. Based on the R^2, welfare weights account for 11–16% of the variation in support across the two measures.[27] We find similarly strong relationships when we use binary versions of the outcomes (split at each outcome's midpoint) and ν (split at 0), presented in Columns (2) and (4).

Table 7: Correlation with Support for Redistribution

	General redistribution		Redistribute at margin		
	(1)	(2)	(3)	(4)	(5)
ν	-0.884***		-0.370***		-0.349***
	(0.051)		(0.022)		(0.023)
$1(\nu > 0)$		0.369***		0.443***	
		(0.026)		(0.025)	
Fair distribution of income					-0.316***
					(0.060)
Constant	4.550***	0.371***	0.967***	0.399***	1.031***
	(0.052)	(0.023)	(0.023)	(0.023)	(0.026)
Observations	1996	1996	1996	1996	1996
R^2	0.158	0.104	0.112	0.177	0.126
Outcome var	Continous	Binary	Continous	Binary	Continous
Explanatory var	Continous	Binary	Continous	Binary	Binary

Notes: The table presents coefficient estimates from linear regressions. The dependent variable in Columns (1) and (2) takes values from 1 through 7, with higher values indicating stronger support for the government to do something to reduce income differences between the rich and poor. The dependent variable in Column (3), (4), and (5) takes values from -2 to $+2$, where positive (negative) values indicate redistribution from high-income (low/middle-income) individuals to low/middle-income (high-income) individuals. A value of zero indicates a desire for no additional redistribution. ν is the progressivity of Social Architects' welfare weights and $1(\nu > 0)$ is an indicator variable that equals 1 if ν is greater than 0 (progressive weights) and 0 otherwise. *Fair distribution of income* equals 1 if a Social Architect indicated that the current income distribution is "somewhat fair" or "very fair" and 0 otherwise. The regression uses data from Wave 1. HC3 standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

[27]We benchmark the predictive power of welfare weights against their stated political affiliation—an important predictor of support for redistribution (e.g., Stantcheva 2021, Kuziemko et al. 2015). We find that political affiliation (an indicator variable that equals 1 for Republicans and 0 otherwise) explains 3-8% of the variation in support for redistribution across the two measures (see Appendix Table A5). Thus, welfare weights have a stronger predictive power than political affiliation.

Having examined individual-level correlations, we now compare aggregate shares. About 78% of Social Architects assign progressive welfare weights—implying a preference to redistribute at the margin—closely matching the 74% of Social Architects supporting additional progressive government redistribution at the margin.[28]

Overall, the close alignment between welfare weights and support for government redistribution, both in individual-level correlations and in aggregate shares, helps validate our measure of welfare weights. It also suggests that calibrating optimal policy formulas using the elicited welfare weights would likely result in policies with broad public acceptance.

4.7 Do Welfare Weights Capture Fairness Concerns?

We also validate our measure of welfare weights by testing whether it captures fairness concerns, as intended. Building on the previous section—which showed that variation in welfare weights predicts variation in support for government redistribution—we explore whether the policy-relevant variation in welfare weights (the part that predicts policy preferences) is explained by fairness concerns.[29] We elicit Social Architects' fairness concerns by asking them to rate the fairness of the current disposable incomes on a five-point scale from "Very unfair" to "Very fair." This measure likely captures several fairness ideals, including redistribution based on the source of income, but it does not capture some important ones, such as equality of opportunity.

Table 7 presents the results. In Columns (3) and (5), the dependent variable is support for increasing government redistribution at the margin. In Column (3), the coefficient on ν is -0.37: Social Architects with more progressive weights (lower values of ν) express stronger support for increasing redistribution at the margin. In Column (5), we add an indicator for fairness perceptions that equals 1 if a Social Architect views the current income distribution as fair and 0 otherwise. Those who perceive the current distribution as fair have lower support for additional redistribution. Importantly, the coefficient on ν declines from -0.37 in Column (3) to -0.35 in Column (5) once fairness concerns are included, suggesting that fairness concerns partially mediate the effect of ν.

However, fairness concerns do not explain all of the policy-relevant variation in welfare

[28] Support for redistribution at the margin (rather than general support for redistribution) is the right measure for comparison to welfare weights because the former, similar to the welfare weights measure, is measured at the margin and allows for regressive redistribution. We find slightly lower support for our general redistribution measure: 66% indicate levels of redistribution above the mid-point of 4 on the 7-point scale. However, this share rises to 78% if we include Social Architects with a response of 4 as supporting redistribution.

[29] A simple correlation of welfare weights and fairness concerns would uncover whether the total variation in welfare weights (not just the policy-relevant variation) captures fairness concerns. We find that the correlation between our measure of fairness concerns and support for government redistribution at the margin is -0.34 ($p < 0.01$).

weights: the coefficient on ν remains statistically significant in Column (5). This residual likely reflects both the difficulty in obtaining a broad measure of fairness and the role of other concerns. Indeed, we find that beliefs about taxes hurting the economy and views about the scope of the government are also captured by welfare weights (see Appendix Section E.2).

We also measured Social Architects' perceptions about the level of taxes paid by individuals and the share earning low incomes (below $15,000). While Social Architects do have misperceptions along these margins, the policy-relevant variation in welfare weights does not capture these misperceptions.

4.8 Temporal Stability

To examine the temporal stability of welfare weights, we conducted our study across two waves, fielded four weeks apart. Figure 4 presents a scatter plot of ν across the two waves, using the sample of Social Architects with valid responses in both waves. The points cluster around the 45-degree line, highlighting the temporal consistency in the weights. The Pearson correlation coefficient of ν across the waves is 0.55 ($p < 0.01$), which is similar to the estimates found in the literature.[30] We benchmark the temporal stability of welfare weights against the stability of two measures of support for redistribution collected in both waves. The cross-wave correlation for welfare weights lies between the across-wave correlation for general support for redistribution ($\rho = 0.72, p < 0.01$) and support for redistribution at the margin ($\rho = 0.44, p < 0.01$). We also find a high degree of temporal consistency in types: 81% of Social Architects retain the same type classification (progressive or regressive weights) across waves. Among the remaining 19% who have inconsistent type classifications, half transition from regressive weights in Wave 1 to progressive weights in Wave 2, and the other half transition in the other way.

Next, we explore the aggregate temporal stability of welfare weights, an important input to policy evaluation.[31] The weighted median progressivity in Wave 2 is -0.68, which is slightly more progressive than the Wave 1 baseline estimate of -0.63 (see Figure 3). Because both estimates are weighted, differences in observable sample composition across waves are unlikely to explain the gap. Restricting the sample to Social Architects who provided valid responses in both waves yields very similar aggregate progressivity estimates, providing further evidence that the observable sample composition across waves does not drive the dif-

[30] Most evidence on the stability of social preferences comes from studies that track individuals over relatively long horizons. These studies find modest temporal correlations—0.28 in rural Paraguay (Chuang & Schechter 2015), 0.12–0.28 in Vietnam (Carlsson et al. 2014), and 0.39–0.46 in the United States (Fisman et al. 2023). By contrast, work that looks at much shorter intervals has focused on risk and time preferences, which appear considerably more stable, with correlations of 0.55 (Schoemaker & Hershey 1992) and 0.50 (Wölbert & Riedl 2013) observed for risk preferences and 0.64 for time (Dean & Sautmann 2021)

[31] To the extent that aggregate measures are less affected by individual-level measurement, stable aggregates can inform policy even when individual preferences are measured with noise.

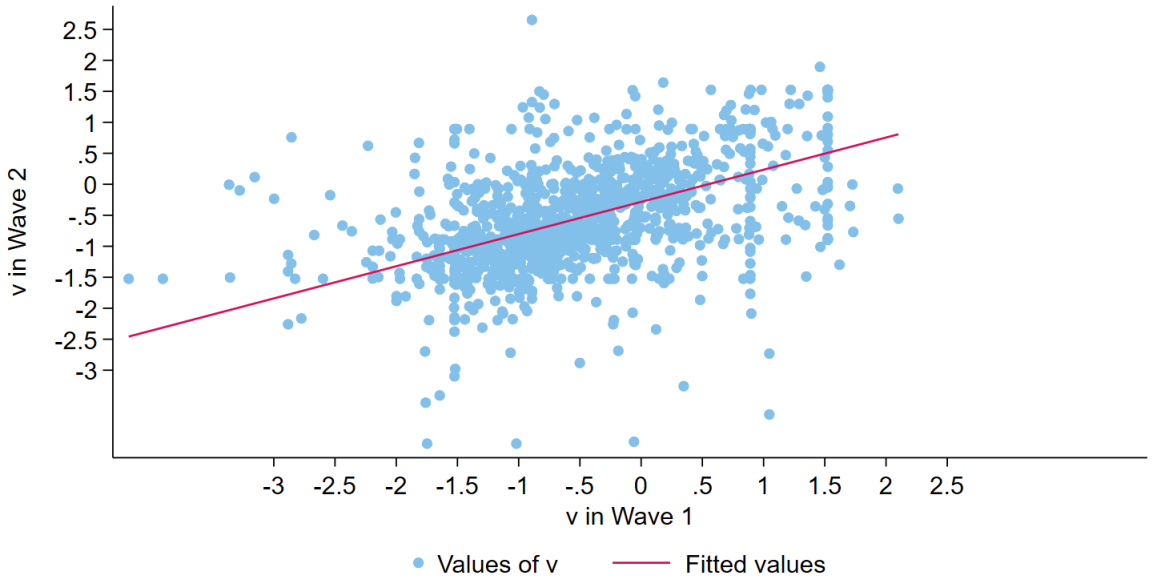


Figure 4: Temporal Stability of Welfare Weights

Notes: The figure presents a scatter plot of the progressivity estimates ν in Wave 1 and Wave 2 using the 1937 sample of Social Architects with valid responses in both waves.

ference in estimates across waves.[32] A more plausible explanation is lower response quality in Wave 1.[33] Consistent with this explanation, the Wave 2 estimate of -0.68 is much closer to the Wave 1 quality-adjusted progressivity estimate of -0.71 (see Section 4.4 for details on quality adjustment). If we additionally adjust responses in Wave 2 for response quality, the progressivity estimate is $\nu = -0.71$, which is even closer to the quality-adjusted wave 1 estimate (see Figure 3).[34] In addition to finding that the median progressivity estimate is stable across waves, we find that the distribution of progressivity estimates is also stable: a Kolmogorov-Smirnov test fails to reject the equality of the distributions of ν across waves ($D = 0.025$, $p = 0.697$).

Overall, we find that welfare weights are temporally stable across waves at the individual and aggregate levels. A high degree of temporal stability at the aggregate level supports using the elicited aggregate progressivity measures to calibrate optimal policy formulas.

[32]The unweighted estimates are: Wave 1: $\nu = -0.66, N = 1397$, bootstrap SE $= 0.03$, 95% normal CI $=$ [–0.72, –0.61]; Wave 2: $\nu = -0.68, N = 1397$, bootstrap SE $= 0.02$, 95% normal CI $=$ [–0.73, –0.64].

[33]The share of Social Architects flagged as having low-quality responses is higher in Wave 1 relative to the Wave 2 for several indicators: failing any comprehension question (7% in Wave 2 and 11% in Wave 1), inconsistent weights (8% in Wave 2 and 11% in Wave 1), and low confidence (34% in Wave 2 and 38% in Wave 1). See Figures A8 and A9 for the shares.

[34]The response quality adjustment in Wave 2 follows the same procedure as that used in Wave 1. Regression estimates of the effect of response quality indicators on the progressivity estimates can be found in Appendix Table A7.

4.9 Summary of Progressivity Estimates

In the previous sections, we found that the weighted median progressivity estimate in Wave 1—our baseline estimate—is given by $\nu = -0.63$. We assessed the robustness of welfare weights to individual-level response quality and corrected for low-quality responses. Since low-quality responses tended to be less progressive, the corrected progressivity estimate of $\nu = -0.71$ is more progressive. We also assessed the robustness of the welfare weights to variation in frames. The progressivity estimates are affected by (i) the first decision, with evidence of anchoring to the first decision, (ii) prompts to consider the consequences of choices, indicating evidence of mistakes, and (iii) the common Recipient across the comparisons, with no evidence of differential response quality across the treatments. We embrace the normative ambiguity across frames that vary the common Recipient, treating both frames as welfare relevant, and additionally, correct these estimates for mistakes based on the prompted treatments. The resulting bounds on the progressivity estimates are given by: $\nu \in [-0.78, -0.70]$. Our response-quality adjusted progressivity estimate of $\nu = -0.71$ also lies within these bounds. Finally, we find that the aggregate progressivity estimates are similar across the two waves of data collection, which supports using the elicited aggregate progressivity measures to calibrate optimal policy formulas.

Comparing our progressivity estimates $\nu \in [-0.78, -0.70]$ to the benchmarks in the literature—which classifies $\nu = |0.25|$ as "weak" redistributive tastes, $\nu = |1|$ as "fairly strong," and $\nu = |4|$ as "extremely strong" (Saez 2002, Allcott et al. 2019)—suggests that our elicited welfare weights fall slightly below the "fairly strong" threshold.

Optimal policy formulas can implement these welfare weights using the parametric function c^{ν} with $\nu \in [-0.78, -0.70]$ or using constant relative risk aversion (CRRA) utilities with a coefficient of relative risk aversion given by $\gamma = -\nu = [0.70, 0.78]$. Our finding of a close alignment between welfare weights and support for government redistribution suggests that calibrating optimal policy formulas with the elicited welfare weights would likely result in policies with broad public acceptance. In Section 5, we compare our estimates of welfare weights to other estimates in the literature, and in Section 6, we highlight the implications of the estimates for optimal income taxes.

Finally, we note that the reforms in our experiment are elicited in a gain-loss frame—one Recipient loses while the other gains—to mirror real policy trade-offs (Saez & Stantcheva 2016, Hendren & Sprung-Keyser 2020). However, we also test whether welfare weights are sensitive to framing in an additional experiment (see Appendix Section F for details and results). We find that welfare weights are slightly more progressive in a gain-gain frame compared to a gain-loss frame.

5 Comparing Welfare Weights

In this section, we compare the welfare weights elicited in our experiment to those implied by the U.S. income tax schedule and by broader tax and transfer policies. The welfare weights implied by policies—referred to as "inverse optimum weights"—represent politicians' aggregation of citizens' welfare weights, possibly influenced by political economy considerations. Details on the estimation of the inverse optimum weights and the comparisons can be found in Appendix Section E.3.

We obtain the inverse optimum weights implied by the U.S. income tax schedule from Hendren (2020). This paper "reveals" the implicit welfare weights from the marginal tax rates (MTRs), the observed income distribution, and estimates of the elasticity of taxable income (ETI). The baseline estimate of inverse-optimum weights uses MTRs in 2012 and mid-range ETI estimates. The resulting weights are decreasing with incomes but increase at the top percentile (see Appendix Figure A10).

The inverse optimum weights implied by U.S. tax and transfer policies are derived using the framework outlined in Hendren & Sprung-Keyser (2020). In this framework, the welfare weight $g(z)$ assigned to beneficiaries of a policy earning z is the inverse of the policy's Marginal Value of Public Funds (MVPF)—the beneficiaries' welfare gain from \$1 of government spending. Intuitively, a planner is willing to incur efficiency losses by implementing a policy with a low MVPF if the beneficiaries are assigned a high welfare weight. We use the MPVF estimates of tax and transfer policies in Hendren & Sprung-Keyser (2020) to infer the welfare weights implied by these policies. The resulting weights are roughly decreasing with the beneficiaries' incomes (See Appendix Figure A11).

We estimate the progressivity of each set of welfare weights using the parametric function c^ν, where c denotes beneficiaries' income and ν is a parameter guiding the progressivity of the weights.[35] The estimated progressivity of the inverse-optimum welfare weights implied by the income tax schedule is $\nu = -0.09$ (SE $= 0.01, 95\%$ CI $= [-0.12, -0.07]$) and for those implied by tax and transfer policies is $\nu = -0.15$ (SE $= 0.08, 95\%$ CI $= [-0.31, -0.00]$). The median welfare weights elicited using our experiment ($\nu \in [-0.78, -0.70]$) are about 8 to 9 times more progressive than the inverse-optimum weights implied by the income tax schedule and about 5 times more progressive than those implied by tax and transfer policies. Figure 5 presents the distribution of welfare weights implied by these estimates against the disposable income distribution.

There are several possible explanations for why the welfare weights we elicit in the experi-

[35]The functional form c^ν, which depends solely on income, can be applied either to pre-tax or post-tax income. Let z denote pre-tax income and $c = h(z)$ the corresponding post-tax income. If welfare weights are based on post-tax income as $g(c)$, then the corresponding weights based on pre-tax income are $g(h(z))$.

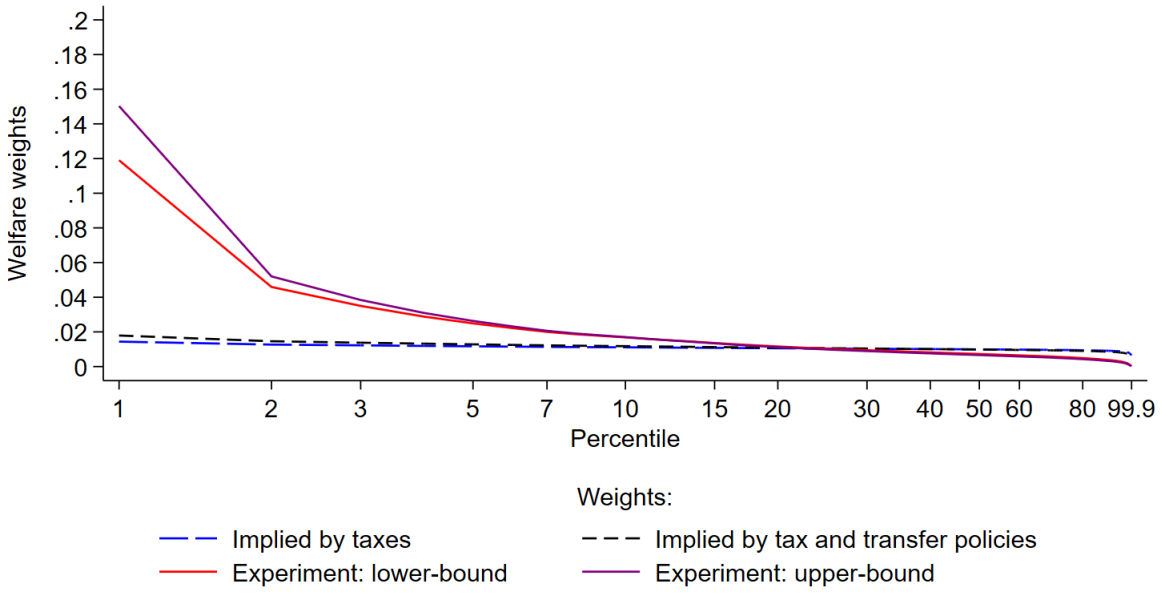


Figure 5: Comparison of Welfare Weights

Notes: The figure plots welfare weights against percentiles of the disposable income distribution. The x-axis has a natural log spacing. Welfare weights are interpolated using c^{ν}, where c is disposable income, and re-normalized to sum to 1. The figure plots the inverse-optimum weights implied by the income tax schedule ($\nu = -0.09$), inverse-optimum weights implied by tax and transfer policies ($\nu = -0.15$), and lower bound ($\nu = -0.70$) and upper bound ($\nu = -0.78$) of the progressivity estimates obtained from the experiment. Data on disposable incomes (which includes in-cash and in-kind transfers) is obtained from Piketty et al. (2018). Individuals with negative disposable income are excluded, and each individual is treated as a single filer.

ment are more progressive than those implied by existing tax and transfer policies. First, political economy considerations—such as politicians aggregating preferences only over certain subgroups (e.g., voters) or overweighting the preferences of high-income individuals—may lead to less progressive inverse-optimum weights. Second, politicians may be aggregating welfare weights that include other considerations, such as self-interest motives, that are absent or less pronounced in our experiment. Finally, differences may arise from assumptions about efficiency costs. In particular, high ETI estimates (than those used by Hendren (2020)) or high perceived efficiency costs by politicians can lead to less progressive inverse-optimum weights.

6 Implications for Optimal Labor Income Taxes

We investigate the implications of the welfare weights elicited in our experiment for optimal non-linear labor income taxes in the U.S. Our calibration applies the optimal tax formula derived by Saez (2001) and adapts the simulation procedures of Mankiw et al. (2009)

and Støstad & Cowell (2024). The optimal tax formula expresses the optimal marginal tax rate (MTR) schedule as a function of (i) the distribution of ability (wages), (ii) the elasticity of taxable income (ETI), and (iii) welfare weights. The calibration relies on the 2019 income distribution obtained from Piketty et al. (2018) to infer the underlying ability distribution, adopts a mid-range ETI of 0.25 (Saez et al. 2012), and uses estimates of welfare weights obtained in our experiment. Additional details are provided in Appendix Section E.4.

Table 8 presents the resulting average optimal MTRs. Using the lower-bound progressivity estimate ($\nu = -0.70$) yields an average optimal MTR of 56%, while the upper-bound estimate ($\nu = -0.78$) yields 58%. For comparison, the average MTR in 2019 was 28%, implying that prevailing rates are roughly 28–30 percentage points below the optimal rates implied by our welfare weights estimates. Closing the gap between existing rates and optimal rates requires more progressive tax rates or higher estimates of the elasticity of taxable income.

Table 8: Average Optimal Marginal Tax Rates (MTRs)

Case	E[MTR]	Guaranteed Income Share
Experiment: lower-bound ($\nu = -0.70$)	56%	57%
Experiment: upper-bound ($\nu = -0.78$)	58%	59%
Current MTRs (2019)	28%	-

Notes: The table presents average optimal MTRs and the guaranteed income transfer to bottom earners (as a share of the average income) calibrated using the lower bound ($\nu = -0.70$) and upper bound ($\nu = -0.78$) of the progressivity estimates obtained from the experiment. The last row presents the average MTRs in the U.S. in 2019. See Appendix Section E.4 for details.

7 Discussion

We elicit the welfare weights assigned by the general population of the U.S. using real-stakes online experiments. In our experiment, participants in the role of Social Architects make several real-stakes redistributive decisions that are used to identify the welfare weights assigned to participants in the role of Recipients. Social Architects' welfare weights are a reduced-form representation of underlying normative ideals, such as equality of opportunity, utilitarianism, redistribution based on the source of income, and poverty alleviation

The aggregate welfare weights are progressive: the median income elasticity of welfare weights is $\nu = -0.63$. We conduct several tests to assess the robustness, validity, and temporal stability of the elicited welfare weights. We assess the robustness of welfare weights to individual response quality and variation in the features of the experimental design. With various

adjustments, we provide bounds on the progressivity estimates, given by $\nu \in [-0.78, -0.70]$. We validate our measure of welfare weights by showing that it correlates with two survey measures of support for government redistribution and captures fairness concerns. Comparing responses across the two waves of data collections, four weeks apart, shows that welfare weights are temporally stable, both at the individual and aggregate levels.

Comparing the estimates of welfare weights obtained from our experiment to the weights implied by policies, we find that the welfare weights elicited in the experiment are about 8 to 9 times more progressive than the inverse-optimum weights implied by the income tax schedule and about 5 times more progressive than those implied by tax and transfer policies. Calibrating optimal income taxes with our estimates of welfare weights shows that the prevailing marginal tax rates are roughly 28–30 percentage points below the optimal rates implied by our welfare weights estimates.

Our reduced-form approach to eliciting welfare weights and the "small-reform" approach to taxation used by Saez & Stantcheva (2016) have a few limitations.

First, the welfare weights estimated in our paper cannot be used to evaluate non-marginal reforms.[36] For non-marginal reforms, the marginal value of the first dollar may differ from the marginal value of the last dollar. We assume a parametric form for the weights—in which welfare weights are a function of consumption—which allows evaluating large reforms, under the assumption that shifts in consumption induced by the reform capture the relevant shifts in Recipient characteristics that Social Architects value.

Second, the welfare weights estimated in our paper can only be used to evaluate policies conditioned on incomes. Future research can explore how welfare weights differ across non-income characteristics, such as age, gender, or ethnicity.

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

References

Abdellaoui, M., Kemel, E., Panin, A. & Vieider, F. M. (2019), 'Measuring time and risk preferences in an integrated framework', *Games and Economic Behavior* **115**, 459–469.

Alesina, A. & Angeletos, G.-M. (2005), 'Fairness and redistribution', *American economic review* **95**(4), 960–980.

[36]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).

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

Allcott, H., Kim, J., Taubinsky, D. & Zinman, J. (2022), 'Are high-interest loans predatory? Theory and evidence from payday lending', *The Review of Economic Studies* **89**(3), 1041–1084.

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.

Allcott, H. & Taubinsky, D. (2015), 'Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market', *American Economic Review* **105**(8), 2501–2538.

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. & Lusardi, A. (2022), 'Evaluating deliberative competence: A simple method with an application to financial choice', *American Economic Review* **112**(11), 3584–3626.

Bernheim, B. D. & Rangel, A. (2009), 'Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics', *The Quarterly Journal of Economics* **124**(1), 51–104.

Berry, J., Fischer, G. & Guiteras, R. (2020), 'Eliciting and utilizing willingness to pay: Evidence from field trials in northern ghana', *Journal of Political Economy* **128**(4), 1436–1473.

Bollinger, C. R. (1998), 'Measurement error in the current population survey: A nonparametric look', *Journal of Labor Economics* **16**(3), 576–594.

Bound, J. & Krueger, A. B. (1991), 'The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?', *Journal of labor economics* **9**(1), 1–24.

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

Burchardi, K. B., De Quidt, J., Gulesci, S., Lerva, B. & Tripodi, S. (2021), 'Testing willingness to pay elicitation mechanisms in the field: Evidence from uganda', *Journal of Development Economics* **152**, 102701.

Bursztyn, L., Handel, B. R., Jimenez, R. & Roth, C. (2023), 'When product markets become collective traps: The case of social media'. NBER Working Paper no. 31771.

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.

Carlsson, F., Johansson-Stenman, O. & Nam, P. K. (2014), 'Social preferences are stable over long periods of time', *Journal of public economics* **117**, 104–114.

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.

Chetty, R., Looney, A. & Kroft, K. (2009), 'Salience and taxation: Theory and evidence', *American Economic Review* **99**(4), 1145–1177.

Chuang, Y. & Schechter, L. (2015), 'Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results', *Journal of Development Economics* **117**, 151–170.

Cohn, A., Jessen, L. J., Klašnja, M. & Smeets, P. (2023), 'Wealthy americans and redistribution: The role of fairness preferences', *Journal of Public Economics* **225**, 104977.

Cruces, G., Perez-Truglia, R. & Tetaz, M. (2013), 'Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment', *Journal of public Economics* **98**, 100–112.

Dean, M. & Sautmann, A. (2021), 'Credit constraints and the measurement of time preferences', *Review of Economics and Statistics* **103**(1), 119–135.

Downs, A. (1957), 'An economic theory of political action in a democracy', *Journal of Political Economy* **65**(2), 135–150.

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

Durante, R., Putterman, L. & Van der Weele, J. (2014), 'Preferences for redistribution and perception of fairness: An experimental study', *Journal of the European Economic Association* **12**(4), 1059–1086.

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

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

Fehr, E., Epper, T. & Senn, J. (2024), 'Social preferences and redistributive politics', *Review of Economics and Statistics* pp. 1–45.

Fehr, E. & Schmidt, K. M. (1999), 'A Theory of Fairness, Competition, and Cooperation', *Quarterly journal of Economics* pp. 817–868.

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. & Vannutelli, S. (2023), 'The distributional preferences of americans, 2013–2016', *Experimental Economics* **26**(4), 727–748.

Fisman, R., Kariv, S. & Markovits, D. (2007), 'Individual preferences for giving', *American Economic Review* **97**(5), 1858–1876.

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

Goldin, J. & Reck, D. (2022), 'Optimal defaults with normative ambiguity', *Review of Economics and Statistics* **104**(1), 17–33.

Hausman, J., Liu, H., Luo, Y. & Palmer, C. (2021), 'Errors in the dependent variable of quantile regression models', *Econometrica* **89**(2), 849–873.

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.

Hvidberg, K. B., Kreiner, C. T. & Stantcheva, S. (2023), 'Social positions and fairness views on inequality', *Review of Economic Studies* **90**(6), 3083–3118.

Jack, B. K., McDermott, K. & Sautmann, A. (2022), 'Multiple price lists for willingness to pay elicitation', *Journal of Development Economics* **159**, 102977.

Jacobs, B., Jongen, E. L. & Zoutman, F. T. (2017), 'Revealed social preferences of dutch political parties', *Journal of Public Economics* **156**, 81–100.

Karadja, M., Mollerstrom, J. & Seim, D. (2017), 'Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution', *Review of Economics and Statistics* **99**(2), 201–212.

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.

List, J. A., Muir, I., Pope, D. & Sun, G. (2023), 'Left-digit bias at lyft', *Review of Economic Studies* **90**(6), 3186–3237.

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.

Luttmer, E. F. P. & Singhal, M. (2011), 'Culture, context, and the taste for redistribution', *American Economic Journal: Economic Policy* **3**(1), 157–179.

Luttmer, E. F. & Samwick, A. A. (2018), 'The welfare cost of perceived policy uncertainty: Evidence from social security', *American Economic Review* **108**(2), 275–307.

Mankiw, N. G., Weinzierl, M. & Yagan, D. (2009), 'Optimal taxation in theory and practice', *Journal of Economic Perspectives* **23**(4), 147–174.

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.

Rees-Jones, A. & Taubinsky, D. (2020), 'Measuring "schmeduling"', *The Review of Economic Studies* **87**(5), 2399–2438.

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.

Schennach, S. (2022), 'Measurement systems', *Journal of Economic Literature* **60**(4), 1223–1263.

Schoemaker, P. J. & Hershey, J. C. (1992), 'Utility measurement: Signal, noise, and bias', *Organizational Behavior and Human Decision Processes* **52**(3), 397–424.

Silva, J. S. & Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and statistics* **88**(4), 641–658.

Singhal, M. (2008), '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* **77**, 48–52.

Stantcheva, S. (2021), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 2309–2369.

Stantcheva, S. (2023), 'How to run surveys: A guide to creating your own identifying variation and revealing the invisible', *Annual Review of Economics* **15**(1), 205–234.

Strulov-Shlain, A. (2023), 'More than a penny's worth: Left-digit bias and firm pricing', *Review of Economic Studies* **90**(5), 2612–2645.

Støstad, M. N. & Cowell, F. (2024), 'Inequality as an externality: Consequences for tax design', *Journal of Public Economics* **235**, 105139.

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

Wölbert, E. & Riedl, A. (2013), 'Measuring time and risk preferences: Reliability, stability, domain specificity'. Working Paper.

Appendix

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

Francesco Capozza

Krishna Srinivasan

Contents

A Variable Definitions

Misperceptions

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

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

- *Gap in top-taxes* = *Beliefs about top-taxes* - 32.7
- *Gap in top-share* = *Beliefs about top-share* - 0.73
- *Gap in non-filers* = 44 - *Beliefs about non-filers*
- *Gap in median-taxes* = *Beliefs about median-taxes* - 13

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

Overestimate share earning $<$ *15K*: We elicit Social Architects' beliefs about the share of households earning less than $15,000. This income level corresponds to the income of the Recipient with the lowest income in our experiment. Social Architects can select a number from 0 to 100 using a slider. We identify Architects' misperceptions by subtracting the actual value (11) from their responses. We obtain the actual value by looking at the share of individuals whose disposable income is below $15,000 in the data obtained from Piketty et al. (2018) (variable *diinc*). Finally, we standardize the misperceptions.

Views about Taxes and Government

We ask Social Architects several questions that elicit their views about the tax system and their trust in government, each capturing a unique mechanism that may explain people's support for redistribution. Many of these questions are drawn from Stantcheva (2021).

Behavioral responses high earners: Takes a value of 1 if a Social Architect indicates that the extent to which taxing high-income earners would encourage them to work less is "A moderate amount," "A lot," or "A great deal," and a value of 0 if the Social Architect indicates "A little," or "None at all."
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."
No Belief trickle down: Takes a value of 1 if a Social Architect indicates that the lower class and working class would "Mostly lose" or "Neither lose nor win" if taxes on high-income earners were cut and a value of 0 if the Social Architect indicates that they would "Mostly win."
Fair distribution of income: Takes a value of 1 if a Social Architect indicates that the current income distribution is "Somewhat fair" or "Very fair" and a value of 0 if the Social Architect indicates "Neither unfair nor fair," "Somewhat unfair," or "Very unfair."
Low trust in government: Takes a value of 1 if a Social Architect trusts the government "Only some of the time" or "Never" and a value of 0 for responses "Most of the time" or "Just about always"
Govt should do less: Takes a value of 1 if a Social Architect indicates thar "Government is doing too much" or Government is doing just the right amount" to solve the country's problems and a value of 0 for "Government should do more."

B Pre-registration

The experimental design, sample restrictions, and analyses were pre-registered. We list the following deviations from the pre-analysis plan:

1. We pre-registered the inclusion of participants with an approval rating between 95 and 100 on Prolific. However, the "representative sample" option on Prolific does not permit screening by approval rate. Matching our data with Prolific's data on participants shows that our sample in Wave 1 includes participants with approval ratings between 93 and 100.

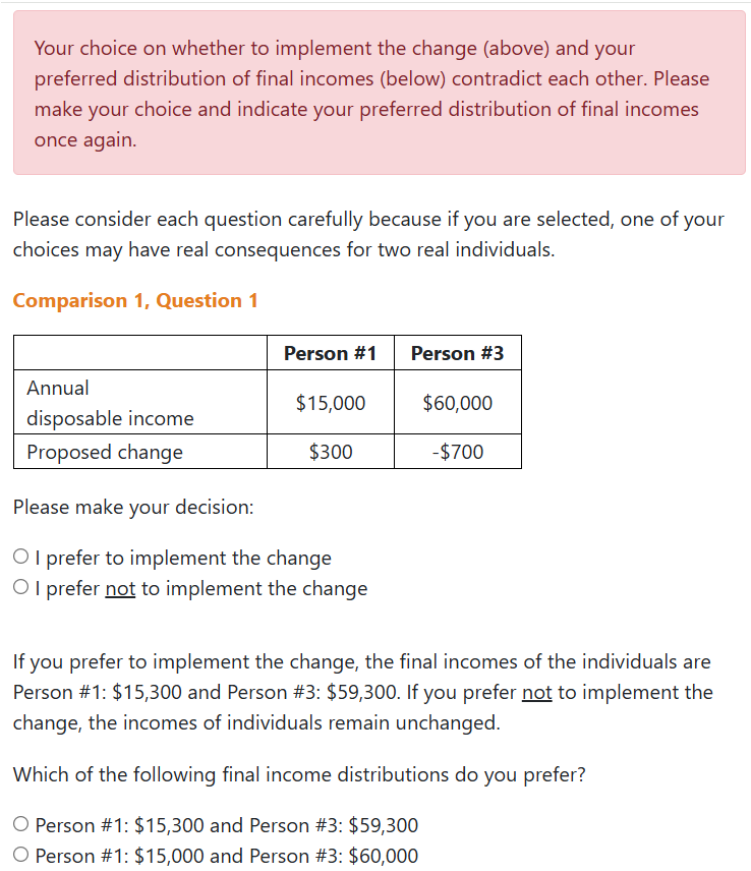
2. For our response-quality adjustment, we pre-registered three separate proxies indicating failure on each of the three comprehension checks. The proportions failing each check are 8%, 3%, and 2%, respectively. To improve power, we instead use a single indicator that flags failure on any of the three checks.

3. We pre-registered flagging respondents with unusual completion times ($\pm$ 2 standard

deviations from the mean) separately by treatment. We create this indicator variable separately for participants in the soft launch (about 250 participants) and the full launch; the median completion time in the soft launch (24.4 minutes) is longer than the completion time in the full launch (14 minutes), because of server congestion during the soft launch.

4. The following exploratory analyses were not pre-registered:
    - Table A5 on the effect of political affiliation on support for redistribution.
    - Table A6 on the effect of the common Recipient on response quality indicators
5. The regression in Column (5) of Table 7 on the mediation analysis was not pre-registered. The pre-registered mediation analysis is discussed in Appendix Section E.2.
6. Following Stantcheva (2021), we reverse code some variables in the mediation analysis discussed in Appendix Section E.2 to ensure that all coefficients have the same sign. This can help understand the overall share explained by the variables.

C Additional Figures



Your choice on whether to implement the change (above) and your preferred distribution of final incomes (below) contradict each other. Please make your choice and indicate your preferred distribution of final incomes once again.

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

Comparison 1, Question 1

	Person #1	Person #3
Annual disposable income	$15,000	$60,000
Proposed change	$300	-$700

Please make your decision:

○ I prefer to implement the change
○ I prefer not to implement the change

If you prefer to implement the change, the final incomes of the individuals are Person #1: $15,300 and Person #3: $59,300. If you prefer not to implement the change, the incomes of individuals remain unchanged.

Which of the following final income distributions do you prefer?

○ Person #1: $15,300 and Person #3: $59,300
○ Person #1: $15,000 and Person #3: $60,000

Figure A1: Screenshot of a Decision Presented to Social Architects in Treatments Prompted

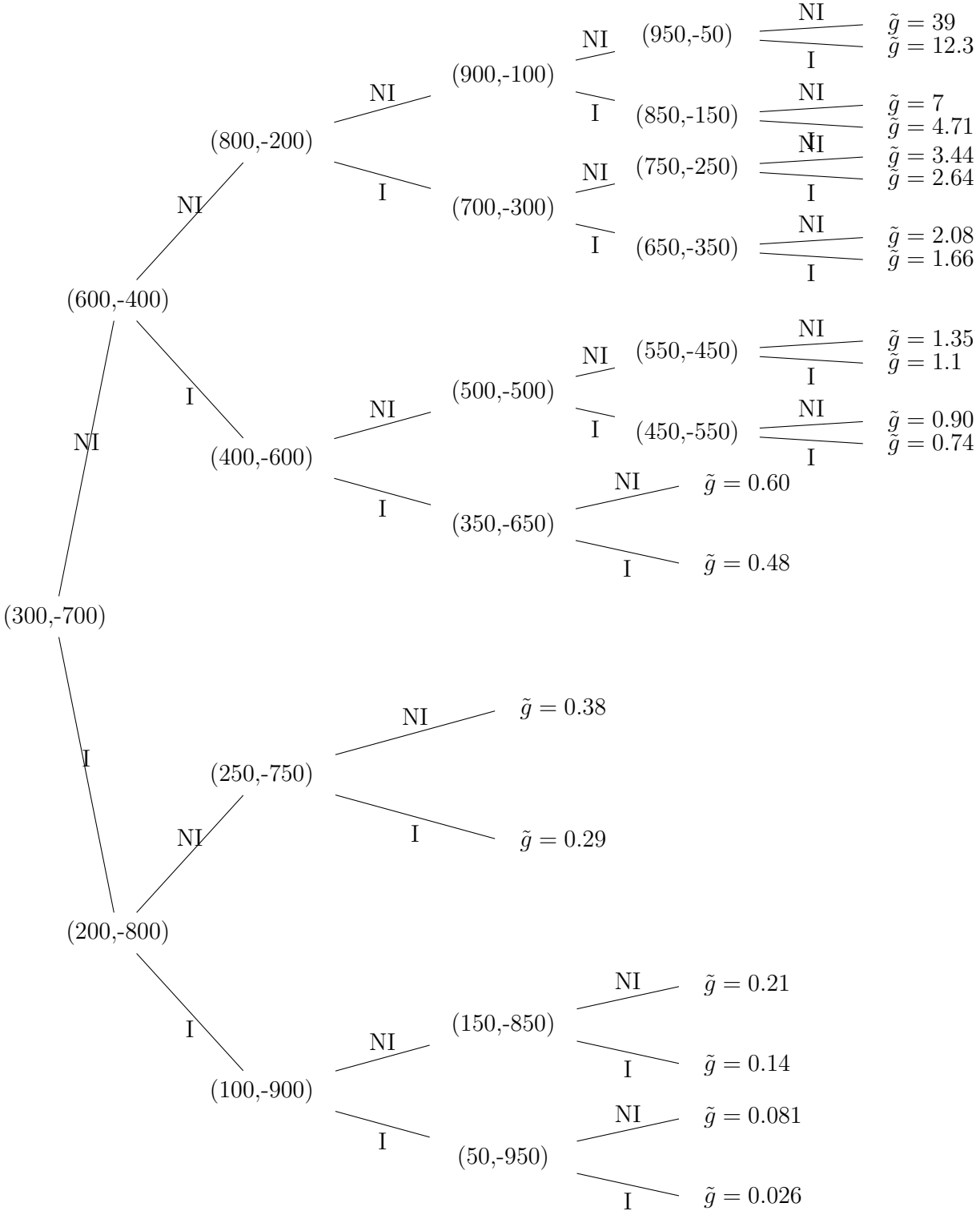


Figure A2: Reforms Selected by the Staircase Method, Starting at (300, −700)

Notes: The figure presents the reforms selected by the staircase method at each node, depending on whether the reform in the previous node was implemented ("I") or not implemented ("NI"). This staircase was implemented in treatments in which the reform in the first decisions is (300, −700). The parameter $\tilde{g}$ is the ratio of the weight assigned to the higher-income Recipient and lower-income Recipient.

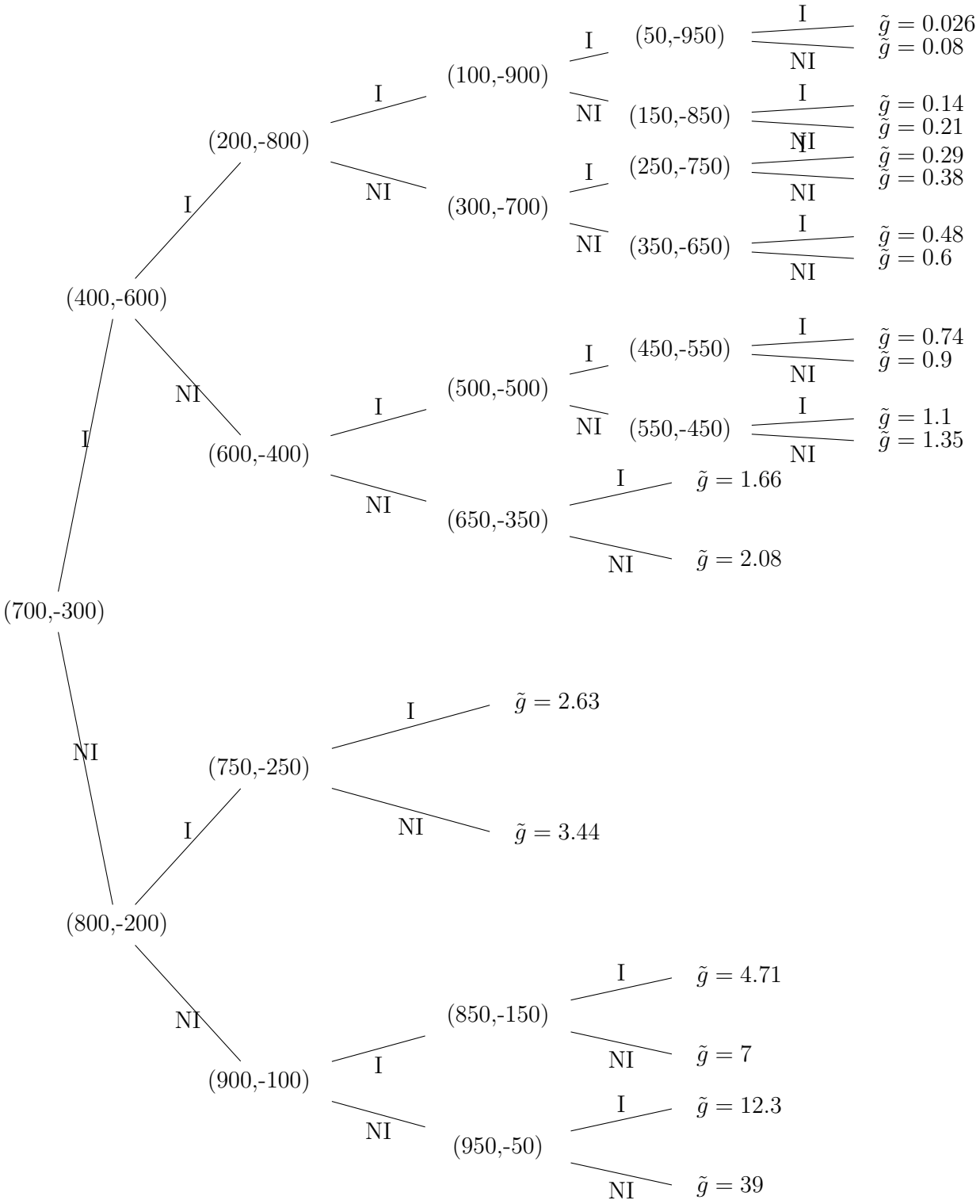


Figure A3: Reforms Selected by the Staircase Method, Starting at (700, −300)

Notes: The figure presents the reforms selected by the staircase method at each node, depending on whether the reform in the previous node was implemented ("I") or not implemented ("NI"). This staircase was implemented in treatments in which the reform in the first decisions is (700, −300). The parameter $\tilde{g}$ is the ratio of the weight assigned to the higher-income Recipient and lower-income Recipient.

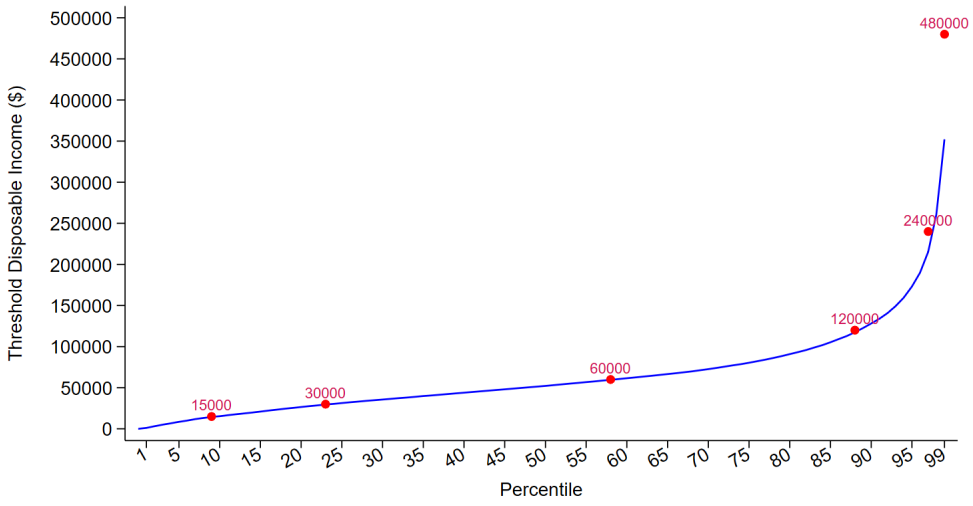


Figure A4: Disposable Incomes of the Six Recipients

Notes: The figure plots the incomes of the six Recipients (dots) against the disposable income distribution (line) in the U.S. in 2019. The horizontal axis indicates the percentiles, and the vertical axis indicates the threshold annual disposable incomes corresponding to the percentiles. Data on disposable incomes (which includes in-cash and in-kind transfers) is obtained from Piketty et al. (2018). Individuals with negative disposable income are excluded, and each individual is treated as a single filer.

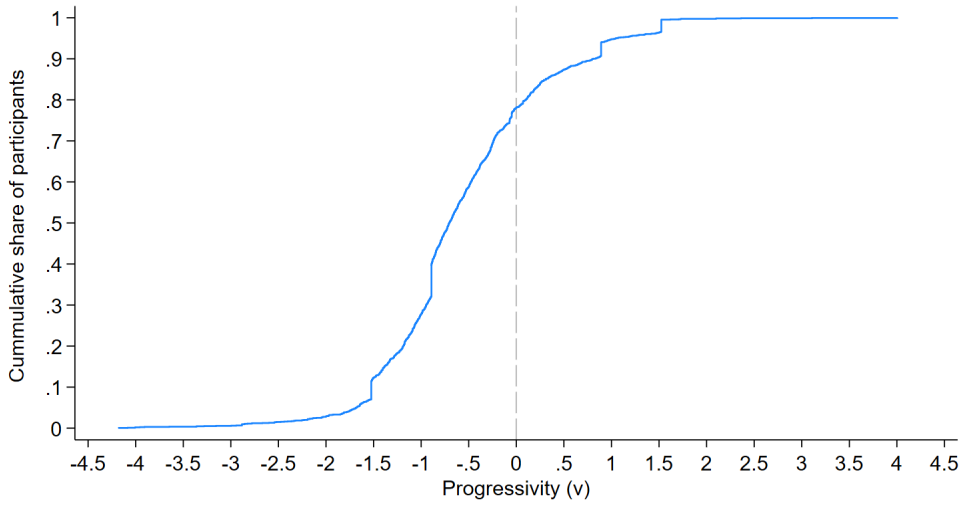


Figure A5: Distribution of Progressivity of Welfare Weights

Notes: The figure presents the cumulative distribution function (CDF) of the progressivity of welfare weights (ν) using data from Wave 1.

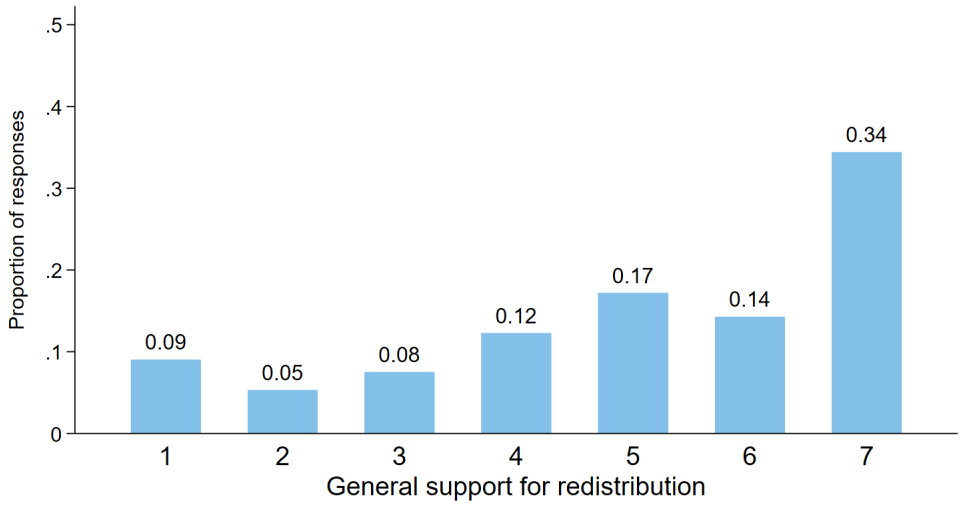


Figure A6: Distribution of General Support for Redistribution

Notes: This figure presents the frequency of responses for the general support for redistribution question. Responses take values from 1 through 7, with higher values indicating stronger support for the government to do something to reduce income differences between the rich and poor. The figure uses data from Wave 1.

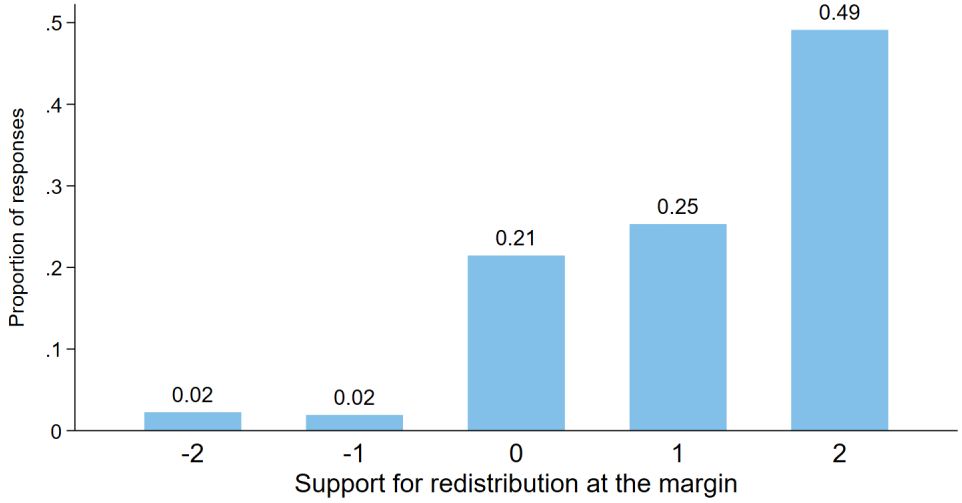


Figure A7: Distribution of Support for Redistribution at the Margin

Notes: This figure presents the frequency of responses for the support for redistribution at the margin question. Responses take values from -2 to $+2$, where positive (negative) values indicate redistribution from high-income (low/middle-income) individuals to low/middle-income (high-income) individuals. A value of zero indicates a desire for no additional redistribution. The figure uses data from Wave 1.

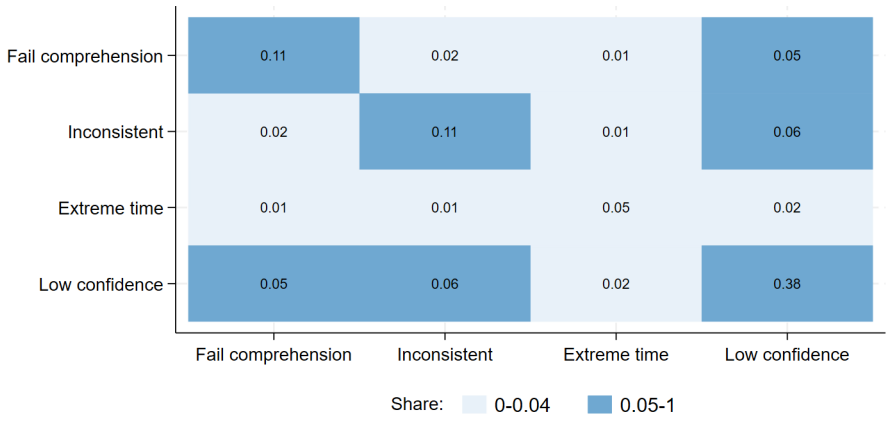


Figure A8: Share Flagged as Low-Quality Based on Indicators in Wave 1

Notes: This figure presents the share of observations where each indicator pair simultaneously equals 1. Diagonals show marginal shares. *Fail comprehension* equals 1 if a Social Architect failed one or more comprehension questions in the first try and 0 otherwise. *Inconsistent* equals 1 if a Social Architect's $\tilde{g}$ in the third and sixth comparisons do not match in sign. *Extreme time* equals 1 if a Social Architect's time spent on the survey lies beyond two standard deviations of the mean. *Low confidence* equals 1 if a Social Architect reports confidence levels lower than the highest category of "Very Much." The figure uses data from Wave 1.

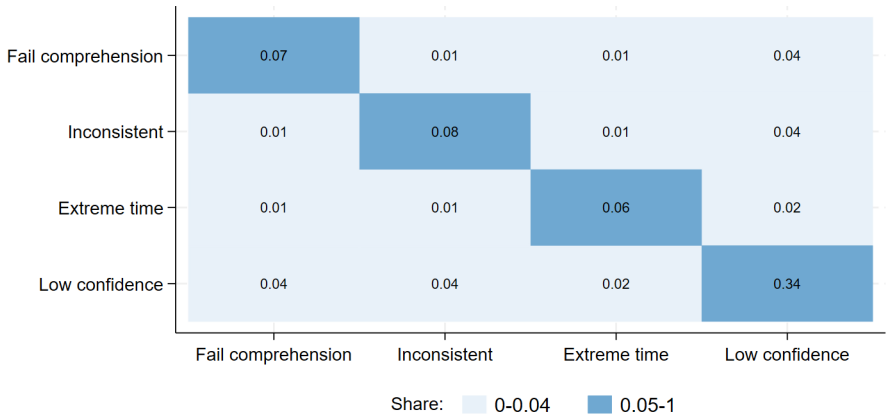


Figure A9: Share Flagged as Low-Quality Based on Indicators in Wave 2

Notes: This figure presents the share of observations where each indicator pair simultaneously equals 1. Diagonals show marginal shares. *Fail comprehension* equals 1 if a Social Architect failed one or more comprehension questions in the first try and 0 otherwise. *Inconsistent* equals 1 if, in the identical third and sixth comparisons, a Social Architect assigns progressive welfare weights ($\tilde{g} < 1$) in one and regressive welfare weights ($\tilde{g} > 1$) in the other. *Extreme time* equals 1 if a Social Architect's time spent on the survey lies beyond two standard deviations of the mean. *Low confidence* equals 1 if a Social Architect reports confidence levels lower than the highest category of "Very Much." The figure uses data from Wave 2.

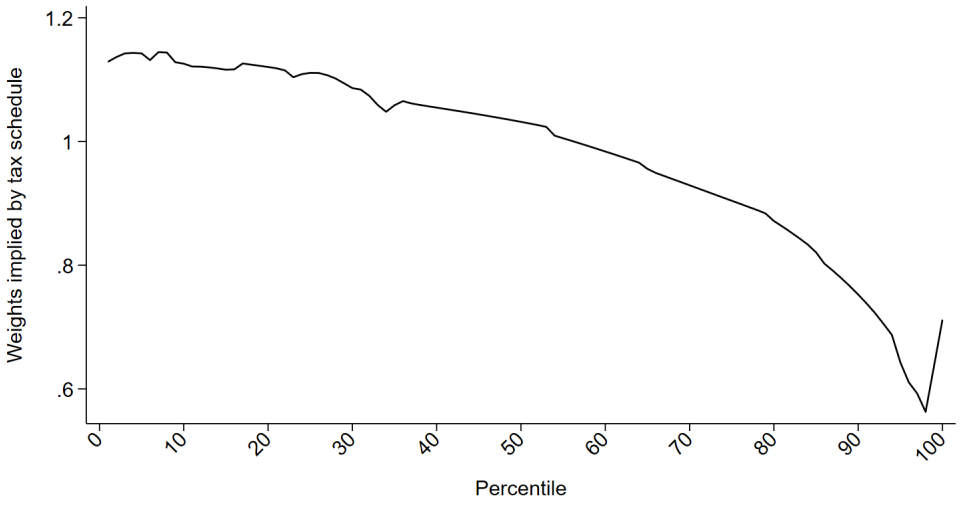


Figure A10: Welfare Weights Implied by the Tax Schedule

Notes: The figure plots the inverse-optimum welfare weights implied by the income tax schedule, obtained from Hendren (2020), for each percentile of the income distribution. We exclude the quintile with negative incomes.

Figure A11: Welfare Weights Implied by Tax and Transfer Policies

Notes: The figure plots the inverse-optimum welfare weights implied by tax and transfer policies against the approximate incomes of the beneficiaries of the policies. The welfare weight of a policy is the inverse of its marginal value of public funds (MVPF). Data on the MVPF of the policies is obtained from Hendren & Sprung-Keyser (2020). See Appendix Section E.3 for details.

D Additional Tables

Table A1: Randomization Check

Variable	p-value Wave 1	Wave 2
Male	0.000	0.000
Income $<$ 30,000	0.536	0.536
Income 30–59,999	0.001	0.001
Income 60–99,999	0.000	0.000
Income 100–149,999	0.005	0.005
Income $\geq$ 150,000	0.000	0.000
Education: High school or less	0.005	0.005
Education: Some college	0.000	0.000
Education: Bachelor or Associate	0.000	0.000
Education: Masters or above	0.094	0.094
Age: 18–24	0.000	0.000
Age: 25–34	0.002	0.002
Age: 35–44	0.000	0.000
Age: 45–54	0.573	0.573
Age: 55–64	0.000	0.000
Age: $\geq$ 65	0.002	0.002
Region: Northeast	0.003	0.003
Region: Midwest	0.000	0.000
Region: South	0.071	0.071
Region: West	0.000	0.000
Republican	0.000	0.000

Notes: The table presents presents the p-values from an F-test from a regression of the particular characteristic on treatment indicators. The second column uses data from Wave 1 and the third column uses data from Wave 2.

Table A2: Example U.S. Tax and Transfer Policies

Policy	**Amount**
Massachusetts one-time electric-bill credit	$50 per residential account
Credit for Other Dependents	Maximum $500
Energy-Efficient Home Improvement Credit	Range from $150 to $1500
Georgia 2025 surplus tax rebate	Maximum range $250-$500
North Dakota Primary Residence Credit	Maximum $1600
NYC Enhanced Real-Property Tax Credit	Maximum $500
Earned Income Tax Credit (no qualifying children, 2024)	Maximum $632

Notes: The table presents several U.S. tax and transfer policies with amounts around $500.

Table A3: Role of Self-Interest Motives

	(1)	(2)
Income near Recipient earning 15K	0.187***	0.199***
	(0.013)	(0.029)
Income near Recipient earning 30K	0.197***	0.106***
	(0.013)	(0.013)
Income near Recipient earning 60K	0.183***	-0.018**
	(0.010)	(0.008)
Income near Recipient earning 120K	0.146***	-0.062***
	(0.013)	(0.014)
Income near Recipient earning 240K	0.166***	-0.027
	(0.036)	(0.043)
Income near earning 480K	0.153*	-0.006
	(0.091)	(0.077)
Constant	0.137***	0.165***
	(0.001)	(0.001)
Observations	11976	11976

Notes: The table presents coefficient estimates from linear regressions. The dependent variable $g(R_j)_i$ is the weight assigned by Social Architect i to Recipient j. In Equation (1), the indicator variables *Income near Recipient earning 15K* through *Income near Recipient 480K* equal 1 if Social Architect i's income lies in each of the following brackets, respectively: [0, $22,500), [$22,500, $45,000), [$45,000, $90,000), [$90,000, $180,000), [$180,000, $360,000), and [$360,000, ∞). In Equation (2), the indicator variables *Income near Recipient earning 15K* through *Income near Recipient 480K* equal 1 if Social Architect i's income is within $\pm$ 20% of the income of the six Recipients, respectively. The regressions use data from Wave 1. Both regressions include Social Architect fixed effects. Robust Standard errors (HC1) in parentheses.
*p<0.1, **p<0.05, ***p<0.01

Table A4: Treatment Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Low-income Recipient on Left	-0.024					-0.044
	(0.047)					(0.045)
Common recipient 60K		-0.138***				-0.131***
		(0.046)				(0.045)
Descending order of income diff			-0.081*			-0.076*
			(0.047)			(0.045)
First decision (300, -700)				-0.171***		-0.188***
				(0.045)		(0.045)
Prompted					-0.066	-0.075*
					(0.047)	(0.045)
Constant	-0.689***	-0.654***	-0.668***	-0.607***	-0.668***	-0.448***
	(0.034)	(0.033)	(0.033)	(0.032)	(0.033)	(0.055)
Observations	1996	1996	1996	1996	1996	1996

Notes: The table presents coefficient estimates from a median regressions. The dependent variable is the progressivity of the welfare weights (ν). The explanatory variables include treatment indicators. The regressions use data from Wave 1. Standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A5: Support for Redistribution and Political Affiliation

	General redistribution	Redistribute at margin
	(1)	(2)
Republican	-1.210***	-0.332***
	(0.097)	(0.048)
Constant	5.428***	1.278***
	(0.048)	(0.025)
Observations	1996	1996
R^2	0.083	0.025

Notes: The table presents coefficient estimates from linear regressions. The dependent variable in Column (1) takes values from 1 through 7, with higher values indicating stronger support for the government to do something to reduce income differences between the rich and poor. The dependent variable in Column (2) takes values from -2 to $+2$, where positive (negative) values indicate redistribution from high-income (low/middle-income) individuals to low/middle-income (high-income) individuals. A value of zero indicates a desire for no additional redistribution. *Republican* equals 1 for Republicans and 0 otherwise. The regression uses data from Wave 1. HC3 standard errors in parentheses.
*p<0.1, **p<0.05, ***p<0.01

Table A6: Quality Indicators and Treatments

	Fail comprehension	Inconsistent	Extreme Time	Low Confidence	Stable
	(1)	(2)	(3)	(4)	(5)
Common recipient 60K	-0.019	0.002	-0.008	0.029	0.033
	(0.014)	(0.014)	(0.010)	(0.022)	(0.021)
Constant	0.120***	0.109***	0.054***	0.368***	0.795***
	(0.010)	(0.010)	(0.007)	(0.015)	(0.015)
Observations	1996	1996	1996	1996	1397

Notes: The table coefficient estimates from linear regressions. *Fail comprehension* equals 1 if a Social Architect failed one or more comprehension questions in the first try and 0 otherwise. *Inconsistent* equals 1 if, in the identical third and sixth comparisons, a Social Architect assigns progressive welfare weights ($\tilde{g} < 1$) in one and regressive welfare weights ($\tilde{g} > 1$) in the other. *Extreme time* equals 1 if a Social Architect's time spent on the survey lies beyond two standard deviations of the mean and 0 otherwise. *Low confidence* equals 1 if a Social Architect reports confidence levels lower than the highest category of "Very Much." *Stable* equals 1 if a Social Architect's welfare weights across the two waves do not match in sign. The explanatory variable is a treatment indicator. The regressions use data from Wave 1. HC3 standard errors in parentheses.
*p<0.1, **p<0.05, ***p<0.01

Table A7: Welfare Weights and Response Quality in Wave 2

	(1)
Fail comprehension	-0.050
	(0.095)
Inconsistent	0.417***
	(0.089)
Extreme time	0.053
	(0.103)
Low confidence	0.089*
	(0.052)
Observations	1397
Controls?	Yes

Notes: The table presents coefficient estimates from a median quantile regression. The dependent variable is the progressivity of the welfare weights (ν). *Fail comprehension* equals 1 if a Social Architect failed one or more comprehension questions in the first try and 0 otherwise. *Inconsistent* equals 1 if, in the identical third and sixth comparisons, a Social Architect assigns progressive welfare weights ($\tilde{g} < 1$) in one and regressive welfare weights ($\tilde{g} > 1$) in the other. *Extreme time* equals 1 if a Social Architect's time spent on the survey lies beyond two standard deviations of the mean. *Low confidence* equals 1 if a Social Architect reports confidence levels lower than the highest category of "Very Much." Controls include treatment indicators (those specified in Table 6) and demographic controls (those specified in Table 4). The regression uses data from Wave 2. Standard errors in parentheses.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

E Additional Analysis

E.1 Correcting for Anchoring

We find that the progressivity of the welfare weights is affected by the first decisions of the staircase method, consistent with Social Architects' anchoring to the first decision. To account for this, we adjust our progressivity estimates following Luttmer & Samwick (2018) and Bursztyn et al. (2023).

Let r_c^r denote the reported reform amount accruing to the low-income Recipient at the point of indifference in comparison c, for $c \in \{1, \ldots, 5\}$. We assume r_c^r is a weighted average of the starting value r^s (either \$300 or \$700, depending on the treatment) and the actual value r_c^a:

$$r_c^r \;=\; \beta_c\, r^s \;+\; (1-\beta_c)\, r_c^a. \tag{A1}$$

By rearranging the terms, we obtain the actual value of the reform:

$$r_c^a \;=\; r_c^r \;-\; \frac{\beta_c}{1-\beta_c}\big(r^s - r_c^r\big). \tag{A2}$$

We estimate β_c by regressing the reported value r_c^r on the starting value r^s, including controls for response quality (described in Section 4.4) and demographics (described in Section 4.3). Table A8 reports the regression results. We find only moderate effects of anchoring, with values of β_c ranging from 0.06 to 0.11. The coefficients in only two of the five regressions are statistically significant at conventional levels ($p < 0.05$).

Using the estimates β_c for each comparison $c \in \{1, \ldots, 5\}$, we compute the adjusted reform amounts using Equation A2. We truncate these adjusted reform amounts at \$25 and \$975—the bounds of the unadjusted reform amounts. We then estimate welfare weights with these adjusted reform amounts using the procedures described in the main text.

E.2 What Do Welfare Weights Capture?

In the main text, we showed that the relationship between welfare weights and support for redistribution is mediated by fairness concerns. Here, we test whether other concerns and misperceptions also mediate the relationship. We proceed in two steps. First, we estimate the overall variation in support for redistribution at the margin that can be explained by welfare weights using a linear regression. Some of this overall variation may be explained by various concerns, which "mediate" the effect of welfare weights. In the second step, we decompose this overall variation in support for redistribution explained by welfare weights into the variation explained by each concern using a Gelbach decomposition (Gelbach 2016).

Table A8: Adjustment of Starting Point

	r_1^r (1)	r_2^r (2)	r_3^r (3)	r_4^r (4)	r_5^r (5)
Starting value 300	0.071**	0.115***	0.066*	0.068*	0.062*
	(0.034)	(0.034)	(0.034)	(0.036)	(0.036)
Observations	1996	1996	1996	1996	1996
Controls?	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficient estimates from linear regressions. The dependent variables in Columns (1)–(5) are the reported reform amounts accruing to the low-income Recipient at the point of indifference in each of the five comparisons: $r_c^r \, \forall \, c \in \{1, \ldots, 5\}$. *Starting value 300* equals 300 if a Social Architect is assigned to a treatment in which the first decision is ($300, −$700) and 700 if the first decision is ($700, −$300). Controls include response quality indicators (described in Section 4.4) and demographic indicators (described in Section 4.3). The regressions use data from Wave 1. HC3 standard errors in parentheses.
*p<0.1, **p<0.05, ***p<0.01

Table A10 presents linear regressions in which the dependent variable is Social Architects' support for redistribution at the margin (taking values from −2 to +2). In Column (1), the explanatory variable is $1(\nu > 0)$, which is an indicator variable taking a value of 1 if a Social Architect has progressive weights and 0 otherwise. The coefficient estimate of $1(\nu > 0)$, is 0.44: this is the overall variation in support for redistribution at the margin explained by welfare weights. In Column (2), we add additional variables to the regression, which capture various concerns and perceptions. Definitions of the variables can be found in Appendix Section A. When we add other variables to the regression, the coefficient estimate of $1(\nu > 0)$ drops to 0.32. This suggests that about 29% of the overall variation in support for redistribution that is explained by welfare weights goes through these other variables.

To identify how much of the overall variance is explained by each of the other variables, we conduct a decomposition procedure proposed by Gelbach (2016). Table A9 presents the results. We find that fairness concerns explain 3% of the overall variation in support for redistribution that is explained by welfare weights; beliefs about higher taxes hurting the economy explain 9% and views about the scope of government explain 14%. Social Architects' misperceptions about the level of taxes paid by individuals and the share of individuals with incomes below $15,000 are including their

We also measured Social Architects' misperceptions about the level of taxes paid by individuals and the share of individuals with incomes below $15,000. While Social Architects do have misperceptions along these margins, their welfare weights do not capture their misperceptions.

Table A9: Decomposition of Treatment Effects

	(1)
Fair distribution of income	3.01
Behavioral responses high-earners	0.83
High taxes high-incomes hurt economy	8.33
No belief trickle down	0.20
Low trust in government	0.09
Govt is doing too little	14.07
Overestimate taxes	0.01
Overestimate low-income share	0.13
Share unexplained	70.94

Notes: The table presents the Gelbach decomposition of the treatment effects following Gelbach (2016). Each value indicates the share of the treatment effects explained by each of the factors, where the dependent variable is the share of policies supported. Share unexplained is the share of the treatment effects not explained by the factors.

Table A10: Support for Redistribution and Mechanisms

	Redistribute at margin	
	(1)	(2)
If Progressive ($\nu > 0$)	0.443***	0.314***
	(0.021)	(0.021)
Fair distribution of income		-0.110***
		(0.024)
Behavioral responses high-earners		-0.069***
		(0.019)
High taxes high-incomes hurt economy		-0.154***
		(0.022)
No belief trickle down		-0.038
		(0.023)
Low trust in government		-0.064***
		(0.021)
Govt Should do less		-0.203***
		(0.019)
Overestimate taxes		0.002
		(0.009)
Overestimate low-income share		0.004
		(0.008)
High Income		-0.018
		(0.017)
High Age		-0.056***
		(0.017)
Low Education		-0.024
		(0.020)
Female		-0.042**
		(0.016)
Republican		-0.044**
		(0.019)
Constant	0.399***	0.815***
	(0.019)	(0.041)
Observations	1996	1996

Notes: The table coefficient estimates from linear regressions. The dependent variable takes values from -2 to $+2$, where positive (negative) values indicate redistribution from high-income (low/middle-income) individuals to low/middle-income (high-income) individuals. A value of zero indicates a desire for no additional redistribution. Definitions of the explanatory variables can be found in Appendix Section A. HC3 standard errors in parentheses.
*p<0.1, **p<0.05, ***p<0.01

E.3 Comparison to Inverse-Optimum Weights

Inverse-Optimum Weights Implied by Tax Schedule: We obtain the inverse optimum welfare weights implied by the U.S. income tax schedule from Hendren (2020). Hendren (2020) computes MTRs using the universe of tax returns in 2012, incorporating ordinary income taxes, the alternative minimum tax (AMT), the earned income tax credits (EITC), and state, local, and Medicare taxes. The baseline estimates of ETI are: top earners (0.30); bottom earners subject to EITC in the phase-in region (0.31) and in the phase-out region (0.14); mid-range estimate elsewhere (0.30). We exclude the bottom quintile with negative incomes from our analysis. Figure A10 plots the inverse-optimum weights implied by the tax schedule against the quintiles of the income distribution.

Inverse-Optimum Weights Implied by Tax Schedule: The inverse optimum weights implied by U.S. tax and transfer policies are derived using the framework outlined in Hendren & Sprung-Keyser (2020). In this framework, the welfare weight $g(z)$ assigned to beneficiaries of a policy earning z is the inverse of the policy's Marginal Value of Public Funds (MVPF)—the beneficiaries' welfare gain from \$1 of government spending. The MVPF of a policy affecting beneficiaries with incomes z^* 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. To replicate this policy (with benefits s^*) through adjustments to the tax schedule, the policy would cost $s^*g(z^*)$, where $g(z^*)$ is the welfare weight. It would be cheaper to achieve s^* through the policy than through adjustments to the tax schedule if and only if $c \leq s^*g(z^*)$. Rewriting this expression yields the following equation: $MVPF = s^*/c \geq 1/g(z^*)$. We focus on tax and in-cash transfer policies, which are the closest analogues to the reform in our experiment: a \$1 spending costs the government roughly \$1, generates benefits of \$1, and is typically conditioned only on income. These include policies in Figure 6 (Panel A) of Hendren & Sprung-Keyser (2020). Additionally, we drop policies with an MVPF of infinity or a negative MVPF, as our theoretical framework and experimental design cannot accommodate such values. Figure A10 plots the inverse-optimum weights implied by the tax and transfer policies against the income distribution.

Estimating the Progressivity of Welfare Weights: We estimate the progressivity of the inverse-optimum welfare weights implied by the tax schedule by regressing the welfare weights on the log of income, and estimating this regression using a Poisson pseudo–maximum likelihood (PPML) estimation. We estimate the progressivity of the inverse-optimum welfare weights implied by tax and transfer policies similarly.

Plotting Welfare Weights: Figure 5 in the main text plots the three sets of welfare weights against the disposable income distribution. The x-axis plots the average disposable income of each percentile. For each set of weights, we interpolate using the function c^{ν}, where c denotes disposable income and ν is the estimated progressivity parameter. Finally, each set of welfare weights is normalized to sum to one. Disposable income (including in-cash and in-kind transfers) is taken from the World Inequality Database (WID). We exclude individuals with negative disposable income and treat each individual as a single filer.

E.4 Calibration of Optimal Income Taxes

We explore the implications of the welfare weights elicited in the experiment for the optimal non-linear labor income taxes in the U.S. We use the optimal tax formula derived in Saez (2001), and follow the simulations presented in Mankiw et al. (2009) and Støstad & Cowell (2024).

The optimal tax formula solves a social planner's problem of maximizing social welfare subject to constraints. It expresses optimal marginal tax rates (MTRs) as a function of (i) the shape of the ability (wage) distribution, (ii) the elasticity of taxable income, and (iii) welfare weights. We estimate the ability distribution directly from the observed U.S. income distribution, apply elasticity estimates from existing literature, and incorporate welfare weights estimated in our paper.

Optimal Tax Formula

We assume that individuals are on a continuum of abilities (wage) w with densities $f(w)$ and cumulative distribution function $F(w)$. Individuals earn income $z = wl$, where l denotes labor supply. Taxes are denoted by $T(z)$, and consumption by $c = z - T(z)$. The planner observes only income z and labor supply l, but not abilities w. Individuals' utilities are given by

$$U(c, l) = u(c) + v(l) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{l^{1+\frac{1}{E_L}}}{(1+\frac{1}{E_L})}, \tag{A3}$$

which are additively separable in consumption and labor. The utility function exhibits constant relative risk aversion (CRRA) in consumption, with the coefficient of relative risk aversion given by γ. Furthermore, it is isoelastic in labor, with E_L denoting the elasticity of earnings with respect to the retention rate $1 - T'(z)$. We denote the utility of an individual with wages w as $U(w)$.

The planner maximizes a lienar utilitarian social welfare function given by $W = \int U(\theta) f(\theta) d\theta$, where θ indexes wages in the integral. Following Saez (2001), the planner's first-order condition is given by

$$\frac{T'(z(w))}{1-T'(z(w))} = \left(\frac{1+E_L^U(w)}{E_L^C(w)}\right)\frac{u'(c(w))}{wf(w)}\int_w^\infty \frac{f(\theta)}{u'(c(\theta))}d\theta - (1-F(w))\frac{1}{p}, \tag{A4}$$

where $T'(z(w))$ is the optimal MTR at income level $z(w)$ at ability level w.

The term $u'(c(w))$ is the marginal utility of consumption or welfare weights. With CRRA utilities, $u'(c(w)) = c(w)^{-\gamma}$. Different estimates of welfare weights can be implemented using different values of γ. The progressivity of the optimal MTRs is increasing with the progressivity of the welfare weights.

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

In Equation (A4), p is the marginal value of public funds. It measures the increase in social welfare obtained when the planner loosens the budget constraint. The cost in consumption terms of a marginal increase in utility for an individual with wage w is $\frac{1}{u'(c(w))}$. The cost of a marginal increase in average utility is $\int_0^\infty \frac{1}{u'(c(w))}f(w)dw$. The value to the planner of a marginal unit of public funds is the inverse of this cost, which is given by

$$p = \frac{1}{\int_0^\infty \frac{1}{u'(c(w))}f(w)dw} \tag{A5}$$

Estimating the Wage-Ability Distribution

We estimate the ability (wage) distribution from the current income distribution and use this exogenous ability distribution when we calibrate the tax formula in Equation (A4). Data on the income distribution is obtained from the Distributional National Accounts micro-files of Piketty et al. (2018). Each observation in the data corresponds to a tax unit.

Step 1: Use the NBER TAXSIM model to find the marginal tax rate for each tax unit. The tax rates are calculated based on the available information about the tax units, which include the number of dependents, the age of the primary filer, and marital status. Add a 5% state tax rate, a 2.9% tax rate for Medicare, and a 2.3% sales tax rate.
Step 2: Assuming that individuals have correctly optimized according to their utility function in Equation (A3), back out the resulting ability (w) of each tax unit.

Step 3: Create a smooth ability distribution using a Kernel density estimator with a bandwidth of \$5000. The smooth distribution has 50,000 observations.
Step 4: Replace the top 0.5% of the distribution with a Pareto distribution. The Pareto parameter is the value of the Pareto parameter $\alpha(z(w)) = \frac{z(w)\cdot f(z(w))}{1-F(z(w))}$ just before the top 0.5%.

Calibrating Optimal Income Taxes

We use an updating rule to find the fixed-point tax schedule, drawing from Mankiw et al. (2009) and Støstad & Cowell (2024). We assume an initial tax schedule. Given the tax schedule, we estimate individuals' labor supply and utilities. Given the utilities, we calculate the resulting optimal MTRs at each wage level using Equation (A4). We iterate on this process until an optimum is found. We check if the second-order condition holds at the optimum. This condition states that the pre-tax incomes are non-decreasing with wages.
Step 1: Start with an initial flat tax rate of 35%.
Step 2: Compute individuals' labor choices based on this tax rate, assuming that they have correctly optimized their utilities in Equation (A3). Computing the derivative of Equation (A3) with respect to l and setting it to 0 yields $l = (w \cdot (1 - T'(z)))^{E_L}$. Set $E_L = 0.25$, which is a mid-range estimate for the elasticity of taxable income (Saez et al. 2012).
Step 3: Based on the optimal labor choices computed in Step 2, calculate the optimal income choices $z = wl$ and the resulting utilities based on Equation (A3). Different values of γ in Equation (A3) lead to different estimates of welfare weights.
Step 4: Calculate the resulting optimal tax rate at each ability level based on Equation (A4).
Step 5: Repeat the previous steps until the tax rates converge to a fixed point.

F Additional Studies

We conducted two additional studies (Study 2 and Study 3), prior to conducting Study 1. All three studies are broadly similar, but with some important differences. First, in Study 1, the Social Architects are asked to decide between implementing a reform (r_l, $-r_h$) or maintaining the status quo (0, 0), while those in the additional studies are asked to decide between implementing a reform (\$500, -\$500) or implementing ($r_l + 500$, $-r_h - 500$). We hypothesize that the simplification of the task in Study 1 likely led to improvements in data quality. However, we are unable to test this because we did not include proxies of response quality in the additional studies. Second, Study 1 has six Recipients earning disposable incomes \$15,000, \$30,000, \$60,000, \$120,000, \$240,000, and \$480,000, while the additional studies have seven Recipients with disposable incomes \$8,000, \$35,000, \$70,000, \$100,000, \$170,000, \$250,000, and \$500,000. Third, Study 1 included proxies of response quality and randomizations, including the randomization of the starting decision and prompts to consider the consequences of choices, which were not included in the additional studies. Overall, Study 1 was designed to be simpler and included several additional checks of robustness. The complete set of instructions for Studies 2 and 3 can be found in Appendix Section H and Section I, respectively.

F.1 Study 2

Data Collection and Sample: In Study 2, we recruited participants in the role of Social Architects from the general population of the U.S. using the data collection provider Lucid. Participants first answer questions about their demographics and political affiliation. We define quotas for recruitment based on gender, age, education, individual income, and region. The quotas are designed to match the sample to the population of the U.S. Participants had to pass one attention check and two comprehension checks to continue with the study. We implemented the survey using Qualtrics. The data collection for Study 2 began on 8 December 2021 and lasted approximately two weeks. Our final sample includes 1965 participants.

Design: The welfare weights elicitation procedure in Study 2 is broadly similar to that in Study 1. In Study 2, Social Architects are randomly assigned to one of four treatments in a 2×2 design. The first treatment dimension tests if Social Architects' welfare weights are sensitive to the framing of the reforms. While Treatments Loss involve taking money away from the higher-income Recipient and giving money to the lower-income Recipient, Treatments Gain involve giving money to both Recipients in the pair. In Treatments Loss, each Recipient is given an initial endowment of \$1500, and a Social Architect decides between

the reforms (r_l, $-r_h$) and (\$500, $-$\$500). In Treatments Gain, the endowment is included in the reform amounts: a Social Architect decides between ($\$1500 + \$r_l, \$1500 - \$r_h$) and ($\$2000, \$1000$). The framing of the reform should not affect a Social Architect's assigned welfare weights since the welfare weights depend on Recipients' consumption, which is the same in the two treatments. The second treatment dimension tests if Social Architects' welfare weights are sensitive to the income of the Recipient common across the decision screens. In Treatments 70K, the Recipient common across the decision screens has an income of \$70,000. In contrast, in Treatments 500K, the Recipient common across the decisions screens has an income of \$500,000.

Results: Column (1) of Table A11 presents estimates from a median regression. The dependent variable is Social Architects' progressivity of the weights (ν). The explanatory variable includes treatment indicators, with the base category being Treatment Loss x 70K. We find that Social Architects have more progressive welfare weights in Treatment Gain $\times$ 70K (ν is 0.12 lower) relative to Treatment Loss $\times$ 70K, and this effect is statistically significant. We do not find an effect across treatments Loss x 500K and Gain x 500K

Table A11: Support for Redistribution and Mechanisms

	(1)	(2)
Gain x 70K	-0.125***	
	(0.038)	
Loss x 500K	0.203***	
	(0.038)	
Gain x 500K	0.203***	
	(0.038)	
Hypothetical		-0.307***
		(0.096)
Constant	-0.203***	-0.185***
	(0.027)	(0.068)
Observations	1965	997
Study:	Study 2	Study 3

Notes: The table presents coefficient estimates from a median regression. The dependent variable is the progressivity of the welfare weights (ν). The explanatory variables include treatment indicators. The regression in Column (1) uses data from Study 2 while the regression in Column (2) uses data from Study 3. Standard errors in parentheses.
$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

F.2 Study 3

Data Collection and Sample In Study 3, we recruited participants in the role of Social Architects from the data collection provider Prolific. The recruitment procedure is similar to the procedure used in Study 2. Participants were recruited based on three quotas available in Prolific: age, sex, and ethnicity. We implemented the survey using oTree (Chen et al. 2016). The data collection for Study 3 began on 14 December 2022 and lasted eight days. Our final sample includes 1992 participants.

Design: The welfare weights elicitation procedure in Study 3 is very similar to that in Study 2. In Study 3, Social Architects are randomly assigned to one of four treatments. The first two treatments test whether Social Architects' welfare weights are sensitive to the existence of real stakes, while the last two treatments focus on the role of self-interest motives. Here, we only discuss the results on the role of stakes. In Treatment Real, Social Architects make real decisions regarding real Recipients, while in Treatment Hypothetical, Social Architects make hypothetical decisions regarding hypothetical Recipients.

Results: Column (2) of Table A11 presents estimates from a median regression. The dependent variable is Social Architects' progressivity of the weights (ν). The explanatory variable is an indicator variable taking a value of 1 if a Social Architect is in Treatment Hypothetical and 0 if the Social Architect is in Treatment Real. We find that the welfare weights are more progressive with hypothetical stakes (ν is 0.30 lower) than with real stakes. This result is consistent with an explanation of Social Architects considering the trade-offs more carefully in the latter.

G Instructions - Study 1

G.1 Wave 1

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

Screen Break

[Consent screen]

Introduction
Welcome! This study is conducted by Unidistance Suisse, Switzerland and WZB, Germany. Our goal is to understand the views of U.S. residents on various topics. By carefully completing this survey, you are helping us to understand these views.

Requirements
To participate in this study, you must be a U.S. resident and at least 18 years old.

Time required
This study will take around **12 minutes**.

Compensation
You will receive **$2.5** for completing the study. The payment will be made in the next few days.

Checks
Our survey includes attention checks to test whether participants take our survey carefully. Additionally, we have implemented measures to ensure that participants do not use AI assistance during the surveys. Participants must complete the survey independently, without the help of AI tools. Participants who fail these checks cannot proceed with the survey, and will be asked to return the survey.

Follow-up study
You may be contacted for a follow-up study approximately four weeks from now. We will notify you of the study via Prolific. Your participation in the follow-up is very important to us.

Confidentiality

Your answers will remain anonymous and will be used for scientific purposes only. Strict confidentiality is guaranteed, and your identity can never be associated with your answers.

Voluntary participation
Participation in this study is voluntary. You may withdraw from the study at any time.

Questions about the survey
If you have questions about this study or your rights, please get in touch with us at krishna.srinivasan@unidistance.ch

Consent
I have received the above information and I am willing to participate in the study.

[Yes; No]

What is your Prolific ID?

Screen Break

[Screen shown if participant does not provide consent]

End of survey

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

Please close this survey and return your submission on Prolific by selecting the "Stop without completing" button.

Screen Break

[Block 1: Background Questions]

What is your gender?

[Female; Male; Non-binary; Prefer not to say]

How old are you?

[18 years old - 24 years old; 25 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.]

What is the highest level of education you have completed?

[Primary education or less; Some high school; High school degree/GED; Some college; 2-year college degree; 4-year college degree; Master's degree; Doctoral degree; Professional degree (e.g., JD, MD, MBA)]

The next question is about your **total individual income**, before taxes, last year. 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, before taxes, last year?

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

——— Screen Break ———

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

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

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

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

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

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

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

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

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

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

[Displayed in all cases]

Could you provide your best guess of what your total individual income, before taxes, last year was?

Screen Break

In politics, as of today, do you consider yourself a Republican, a Democrat or an independent?

[Republican; Democrat; Independent]

Screen Break

[Screen shown if participant does not reside in the U.S.]

End of survey

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

Thank you for your time.

Please close this survey and return your submission on Prolific by selecting the "Stop without completing" button.

Screen Break

[Attention check]

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 "Not at all interested" below:

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

Screen Break

[Screen shown if participant failed the attention check]

End of survey

Unfortunately, you failed the attention check.

For this reason, you cannot continue the study and will not receive a payment.

Please close this survey and return your submission on Prolific by selecting the "Stop without completing" button.

Screen Break

[**Block 2: Eliciting Welfare Weights**]

Instructions

In this section, **you will be asked to decide whether you want to change the incomes of six real individuals in society.** These real individuals will be recruited from the U.S. general population. They are above the age of 18 and are U.S. citizens. As we will explain below, your decisions may have real consequences for two of these individuals.

You will be presented with several questions. In each question, you will be presented with two individuals and learn their disposable incomes. **Disposable income is defined as income after all taxes have been paid and transfers have been received (including federal and state taxes and transfers).**

In each question, you will be presented with a proposed change to the incomes of the individuals, and will be asked to indicate whether or not you prefer to implement this change. **If you prefer to implement the change, the income of the lower-income individual in the pair will increase, and the income of the higher-income individual in the pair will decrease by the amounts proposed in the change.** If you prefer not to implement the change, the incomes of the two individuals will remain unchanged. We will describe below how these changes are implemented.

Example

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

	Person #3	**Person #6**
Annual disposable income	\$60,000	\$120,000
Proposed change	\$500	-\$500

Please make your decision:

- ◦ I prefer to implement the change
- ◦ I prefer not to implement the change

In this example, there are two individuals: Person #3 with an income of \$60,000 and Person #6 with an income of \$120,000.

The proposed change involves increasing the income of Person #3 by $500 and decreasing the income of Person #6 by $500. If you prefer to implement the change, the final incomes of the individuals are Person #3: $60,500 and Person #6: $119,500. If you prefer not to implement the change, the incomes of the individuals remain unchanged.

Comparisons

You will face several questions like the one above, with the amounts in the proposed change varying across the questions. You will face several comparisons, with the income of the individuals varying across them. The following table indicates the comparisons.

Comparison	Annual disposable income of individuals
1	$15,000 vs. $60,000
2	$30,000 vs. $60,000
3	$60,000 vs. $120,000
4	$60,000 vs. $240,000
5	$60,000 vs. $480,000

Incentives and Implementation

At the end of the study, one participant will be randomly selected. If you are selected, one question will be randomly chosen from either this survey or the follow-up survey, and your choice on that question will be implemented. The two individuals involved in the selected question will then be affected by your decision. **Thus, if you are randomly selected, one of your choices may have real consequences for two other individuals.**

If you are selected, a $1,000 bonus will be transferred to the two individuals affected by your choice. The disposable incomes shown above for each person already include this $1,000 bonus, under the assumption that they receive it. The amounts specified in your proposed change will be added to or subtracted from that bonus.

[If Treatment Prompted] For the first question in each comparison, you will be presented with two distributions of income for the pairs of individuals, which reflect the consequences of the two choices you were presented with. You will then be asked to decide which distribution you

prefer. This is meant to help you consider the consequences of your choice before proceeding.

Please answer the following questions to show that you have understood the instructions. You can read the instructions above again if needed.

How many individuals will you make decisions regarding?

[Three; Six]

We will present you with the incomes of several individuals. What type of income will we present to you?

[Pre-tax income, Disposable income]

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

[True; False]

Screen Break

If a participant fails any of the three comprehension checks, they are taken back to the instruction screen and informed which checks they failed. The error message reads: "In your first try, you answered the [] question[s] incorrectly. Please try again." Any check answered correctly is locked, while those failed require a new response. Participants must pass the remaining checks before proceeding. On subsequent attempts, if a participant fails a check, they receive the following error message: "In your previous try, you answered the [] question[s] incorrectly. Please try again."

Screen Break

[We present the proposed change in each question as ($\$r_l$, $-\$r_h$), where $\$r_l$ and $-\$r_h$ refer to the amounts accrued to the lower-income Recipient and high-income Recipients in the pair, respectively. Participants are randomized into one of two treatments that vary the amounts in the first question to be either ($300, −$700) or ($700, −$300). Below we describe the questions in both treatments.]

[C1Q1: Comparison 1 Question]

Comparison 1/5, Question 1

[If first decision is ($300, −$700)]

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

	Person #1	**Person #3**
Annual disposable income	$15,000	$60,000
Proposed change	$300	-$700

Please make your decision:

- ◦ I prefer to implement the change
- ◦ I prefer not to implement the change

If you prefer to implement the change, the final incomes of the individuals are Person #1: $15,300 and Person #3: $59,300. If you prefer not to implement the change, the incomes of individuals remain unchanged.

[If first decision is ($700, −$300)]

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

	Person #1	**Person #3**
Annual disposable income	$15,000	$60,000
Proposed change	$700	-$300

Please make your decision:

- ◦ I prefer to implement the change
- ◦ I prefer not to implement the change

If you prefer to implement the change, the final incomes of the individuals are Person #1: $15,700 and Person #3: $59,700. If you prefer not to implement the change, the incomes of individuals remain unchanged.

Screen Break

[If first decision is ($300, −$700)]

[C1Q2.1: If change implemented in C1Q1, choose whether to implement ($200, −$800)]

[C1Q2.2: If change not implemented in C1Q1, choose whether to implement ($600, −$400)]

[If first decision is ($700, −$300)]

[C1Q2.1: If change implemented in C1Q1, choose whether to implement ($400, −$600)]

[C1Q2.2: If change not implemented in C1Q1, choose whether to implement ($800, −$200)]

Screen Break

[If first decision is ($300, −$700)]

[C1Q3.1: If change implemented in C1Q2.1, choose whether to implement ($100, −$900)]

[C1Q3.2: If change not implemented in C1Q2.1, choose whether to implement ($250, −$750)]

[C1Q3.3: If change implemented in C1Q2.2, choose whether to implement ($400, −$600)]

[C1Q3.4: If change not implemented in C1Q2.2, choose whether to implement ($800, −$200)]

[If first decision is ($700, −$300)]

[C1Q3.1: If change implemented in C1Q2.1, choose whether to implement ($200, −$800)]

[C1Q3.2: If change not implemented in C1Q2.1, choose whether to implement ($600, −$400)]

[C1Q3.3: If change implemented in C1Q2.2, choose whether to implement ($750, −$250)]

[C1Q3.4: If change not implemented in C1Q2.2, choose whether to implement ($900, −$100)]

Screen Break

[If first decision is ($300, −$700)]

[C1Q4.1: If change implemented in C1Q3.1, choose whether to implement ($50, −$950)]

[C1Q4.2: If change not implemented in C1Q3.1, choose whether to implement ($150, −$850)]

[If change implemented in C1Q3.2, participant indifferent between ($225, −$775)]

[If change not implemented in C1Q3.2, participant indifferent between ($275, −$725)]

[C1Q4.3: If change implemented in C1Q3.3, choose whether to implement ($350, −$650)]

[C1Q4.4: If change not implemented in C1Q3.3, choose whether to implement ($500, −$500)]

[C1Q4.5: If change implemented in C1Q3.4, choose whether to implement ($700, −$300)]

[C1Q4.6: If change not implemented in C1Q3.4, choose whether to implement ($900, −$100)]

[If first decision is ($700, −$300)]

[C1Q4.1: If change implemented in C1Q3.1, choose whether to implement ($100, −$900)]

[C1Q4.2: If change not implemented in C1Q3.1, choose whether to implement ($300, −$700)]

[C1Q4.3: If change implemented in C1Q3.2, choose whether to implement ($500, −$500)]

[C1Q4.4: If change not implemented in C1Q3.2, choose whether to implement ($650, −$350)]

[If change implemented in C1Q3.3, participant indifferent between ($725, −$275)]

[If change not implemented in C1Q3.3, participant indifferent between ($775, −$225)]

[C1Q4.5: If change implemented in C1Q3.4, choose whether to implement ($850, −$150)]

[C1Q4.6: If change not implemented in C1Q3.4, choose whether to implement ($950, −$50)]

Screen Break

[If first decision is ($300, −$700)]

[If change implemented in C1Q4.1, participant indifferent between ($25, −$975)]

[If change not implemented in C1Q4.1, participant indifferent between ($75, −$925)]

[If change implemented in C1Q4.2, participant indifferent between ($125, −$875)]

[If change not implemented in C1Q4.2, participant indifferent between ($175, −$825)]

[If change implemented in C1Q4.3, participant indifferent between ($325, −$675)]

[If change not implemented in C1Q4.3, participant indifferent between ($375, −$625)]

[C1Q5.1: If change implemented in C1Q4.4, choose whether to implement ($450, −$550)]

[C1Q5.2: If change not implemented in C1Q4.4, choose whether to implement ($550, −$450)]

[C1Q5.3: If change implemented in C1Q4.5, choose whether to implement ($650, −$350)]

[C1Q5.4: If change not implemented in C1Q4.5, choose whether to implement ($750, −$250)]

[C1Q5.5: If change implemented in C1Q4.6, choose whether to implement ($850, −$150)]

[C1Q5.6: If change not implemented in C1Q4.6, choose whether to implement ($950, −$50)]

[If first decision is ($700, −$300)]

[C1Q5.1: If change implemented in C1Q4.1, choose whether to implement ($50, −$950)]

[C1Q5.2: If change not implemented in C1Q4.1, choose whether to implement ($150, −$850)]

[C1Q5.3: If change implemented in C1Q4.2, choose whether to implement ($250, −$750)]

[C1Q5.4: If change not implemented in C1Q4.2, choose whether to implement ($350, −$650)]

[C1Q5.5: If change implemented in C1Q4.3, choose whether to implement ($450, −$550)]

[C1Q5.6: If change not implemented in C1Q4.3, choose whether to implement ($550, −$450)]

[If change implemented in C1Q4.4, participant indifferent between ($625, −$375)]

[If change not implemented in C1Q4.4, participant indifferent between ($675, −$325)]

[If change implemented in C1Q4.5, participant indifferent between ($825, −$175)]

[If change not implemented in C1Q4.5, participant indifferent between ($875, −$125)]

[If change implemented in C1Q4.6, participant indifferent between ($925, −$75)]

[If change not implemented in C1Q4.6, participant indifferent between ($975, −$25)]

Screen Break

[If first decision is ($300, −$700)]

[If Yes chosen in C1Q5.1, participant indifferent between ($425, −$575)]

[If No chosen in C1Q5.1, participant indifferent between ($475, −$525)]

[If Yes chosen in C1Q5.2, participant indifferent between ($525, −$475)]

[If No chosen in C1Q5.2, participant indifferent between ($575, −$425)]

[If Yes chosen in C1Q5.3, participant indifferent between ($625, −$375)]

[If No chosen in C1Q5.3, participant indifferent between ($675, −$325)]

[If Yes chosen in C1Q5.4, participant indifferent between ($725, −$275)]

[If No chosen in C1Q5.4, participant indifferent between ($775, −$225)]

[If change implemented in C1Q5.5, participant indifferent between ($825, −$175)]

[If No chosen in C1Q5.5, participant indifferent between ($875, −$125)]

[If Yes chosen in C1Q5.6, participant indifferent between ($925, −$75)]

[If No chosen in C1Q5.6, participant indifferent between ($975, −$25)]

[If first decision is ($700, −$300)]

[If Yes chosen in C1Q5.1, participant indifferent between ($25, −$975)]

[If No chosen in C1Q5.1, participant indifferent between ($75, −$925)]

[If Yes chosen in C1Q5.2, participant indifferent between ($125, −$875)]

[If No chosen in C1Q5.2, participant indifferent between ($175, −$825)]

[If change implemented in C1Q5.3, participant indifferent between ($225, −$775)]

[If change not implemented in C1Q5.3, participant indifferent between ($275, −$725)]

[If change implemented in C1Q5.4, participant indifferent between ($325, −$675)]

[If change not implemented in C1Q5.4, participant indifferent between ($375, −$625)]

[If change implemented in C1Q5.5, participant indifferent between ($425, −$575)]

[If change not implemented in C1Q5.5, participant indifferent between ($475, −$525)]

[If change implemented in C1Q5.6, participant indifferent between ($525, −$475)]

[If change not implemented in C1Q5.6, participant indifferent between ($575, −$425)]

Screen Break

[Comparisons 2-5 are identical to Decision Screen 1, with the below exceptions]

[In Comparison 2, the Recipients are Person #2: $30,000 and Person #3: $60,000]

[In Comparison 3, the Recipients are Person #3: $60,000 and Person #4: $120,000]

[In Comparison 4, the Recipients are Person #3: $60,000 and Person #5: $240,000]

[In Comparison 5, the Recipients are Person #3: $60,000 and Person #6: $480,000]

Screen Break

[Consistency check]

In this final comparison, all participants will see a pair of individuals that they have encountered before. It is very important for us that you once again consider each question carefully. Thank you very much.

[In this Decision Screen, the Recipients are Person #3: $60,000 and Person #4: $120,000]

Screen Break

[Treatments]

[Participants are randomized to various treatments. The first treatment dimension varies the order of the Comparisons. For half the participants, the order is as presented above, while

for the other half, the order of the first five Comparisons is reversed. The sixth Comparison is identical across the two treatments. The second treatment dimension varies whether the Recipient common across the Comparisons earns an income of \$60,000 (as shown above) or \$120,000. The third treatment dimension varies whether the low-income Recipient is presented on the left (as shown above) or the right of the screen.

The fourth treatment dimension varies whether participants are prompted to consider the consequences of their decisions prior to proceeding. Participants are asked to choose between two final income distributions that reformulate the consequences of their decision. Figure A1 presents a screenshot. If their preference on whether to implement the reform and their preferred final income distribution are consistent, they can proceed with the survey. If there is an inconsistency, they see the following error message: "Your choice on whether to implement the change (above) and your preferred distribution of final incomes (below) contradict each other. Please make your choice and indicate your preferred distribution of final incomes once again.'

The fifth treatment varies the questions in the staircase (explained above), with the reform in the first decision being either (\$300, −\$700) or (\$700, −\$300). The first four treatment dimensions have implications for the instructions and decisions, while the fifth dimension only has implications for the decisions.]

Screen Break

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

[Not at all; Very little; Little; Somewhat; Very much]

Screen Break

[**Block 3: Support for Redistribution**]

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

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.

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; 2; 3; 4; 5:; 6:; 7: Government should not concern itself with income differences]

Screen Break

Consider the current **disposable incomes** of individuals in society. Disposable income is defined as **income after all taxes have been paid and transfers have been received.**

Do you think that, given the current disposable incomes of individuals in society, incomes should be further redistributed or should they remain as they are?

Please provide your answer on a scale from -2 to $+2$.

- A -2 means that income should be further redistributed by taking from lower/middle-income individuals and giving to higher-income individuals.
- A $+2$ means that income should be further redistributed by taking from higher-income individuals and giving to lower/middle-income individuals.

Given the current disposable incomes of individuals in society ...

[-2: Incomes should be further redistributed by taking from lower/middle-income individuals and giving to higher-income individuals; -1:; +0: Incomes should remain as they are; +1:; +2: Income should be further redistributed by taking from higher-income individuals and giving to lower/middle-income individuals].

Screen Break

[**Block 4: Knowledge**]

The next set of questions is about the income tax system in the United States. In order for your answers to be most helpful to us, it is really important that you provide your best guesses to these questions. Although you may find some questions difficult, it is very important for our research that you try your best. Thank you very much!

In 2024, individuals (single filers) with income over $609,350 were in the top federal personal income tax bracket.

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

[slider 0-100]

What is the marginal income tax rate applied to incomes at 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 every 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 a disposable income below $15,000?

[slider 0-100]

Screen Break

[**Block 5: Mechanisms**]

[The order of the questions is randomized across participants]

Please answer the following last set of questions.

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]

Consider the current disposable incomes of individuals in society, defined as income after all taxes have been paid and transfers have been received. Do you think that the current distribution of disposable incomes in society is unfair or fair?

[Very unfair; Somewhat unfair; Neither unfair nor fair; Somewhat fair; Very fair]

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

[Never; Only some of the time; Most of the time; Just about always]

Some people think the government is trying to do too many things that should be left to individuals and businesses. Others think that the government should do more to solve our country's problems. Which comes closer to your own view?

[Government is doing too much; Government is doing just the right amount; Government should do more]

Screen Break

End of survey

Thank you for your time!

We will pay you your $2.5 participation payment in the following days.

Please click the following link to finish the survey:

[link]

G.2 Wave 2

[Consent screen]

Introduction
Welcome! This study is conducted by Unidistance, Switzerland and WZB, Germany. You previously participated in our survey. We invite you to participate in our follow-up survey.

Time required
This study will take around **5 minutes**.

Compensation
You will receive **$1** for completing the study. The payment will be made in the next few days.

Checks
Our survey includes attention checks to test whether participants take our survey carefully. Additionally, we have implemented measures to ensure that participants do not use AI assistance during the surveys. Participants must complete the survey independently, without the help of AI tools. Participants who fail these checks cannot proceed with the survey, and will be asked to return the survey.

Confidentiality
Your answers will remain anonymous and will be used for scientific purposes only. Strict confidentiality is guaranteed, and your identity can never be associated with your answers.

Voluntary participation
Participation in this study is voluntary. You may withdraw from the study at any time.

Questions about the survey
If you have questions about this study or your rights, please get in touch with us at krishna.srinivasan@unidistance.ch

Consent
I have received the above information and I am willing to participate in the study.

[Yes; No]

What is your Prolific ID?

Screen Break

[Screen shown if participant does not provide consent]

End of survey

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

Please close this survey and return your submission on Prolific by selecting the "Stop without completing" button.

Screen Break

[Attention check]

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 "Not at all interested" below:

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

Screen Break

[Screen shown if participant failed the attention check]

End of survey

Unfortunately, you failed the attention check.

For this reason, you cannot continue the study and will not receive a payment.

Please close this survey and return your submission on Prolific by selecting the "Stop without completing" button.

Screen Break

In the next part, we will present you with several decisions. **These decisions may look very similar to the ones you faced in our previous survey.**

However, it is very important for us that you consider these questions carefully.

Screen Break

[**Block 1: Eliciting Welfare Weights**]

[Participants are presented with questions designed to elicit their welfare weights. The questions are identical to those in Wave 1. Each participant is assigned to the same treatment group in both waves. Only one sentence in the instructions for Wave 2 differs from those in Wave 1. The sentence in Wave 1 "If you are selected, we will randomly select one question (from this survey or the follow-up survey) and implement your choice on this question." is replaced in Wave 2 with "If you are selected, we will randomly select one question (from this survey or the previous survey) and implement your choice on this question."]

Screen Break

[**Block 2: Support for Redistribution**]

[Participants are presented with both questions about preferences for redistribution.]

H Instructions - Study 2

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

H.1 Treatment Loss x 70K

[Consent screen]

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

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

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

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

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

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

Consent

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

[Yes; No]

Screen Break

[Screen shown if participant does 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.

Screen Break

[Block 1: Background Questions]

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?

Screen 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.

Screen Break

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

End of survey

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

Thank you for your time.

You will be automatically redirected in 5 seconds.

Screen Break

[Attention check]

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]

Screen Break

[Screen shown 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.

Screen Break

[Block 2: Eliciting Welfare Weights]

[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)

Screen Break

[If a participant fails the comprehension check]

End of survey

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

You cannot continue with the study.

Thank you for your time.

You will be automatically redirected in 5 seconds.

Screen Break

[C1Q1: Comparison 1 Question 1]

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

Screen Break

[All questions hereafter in Decision Screen 1 look like C1Q1]

[C1Q2.1: If $(500, -500)$ chosen in C1Q1, choose between $(1250, -750)$ and $(500, -500)$]

[C1Q2.2: If $(1000, -1000)$ chosen in C1Q1, choose between $(750, -1250)$ and $(500, -500)$]

Screen Break

[C1Q3.1: If $(500, -500)$ chosen in C1Q2.1, choose between $(1375, -625)$ and $(500, -500)$]

[C1Q3.2: If $(1250, -750)$ chosen in C1Q2.1, choose between $(1125, -875)$ and $(500, -500)$]

[C1Q3.3: If $(500, -500)$ chosen in C1Q2.2, choose between $(875, -1125)$ and $(500, -500)$]

[C1Q3.4: If $(750, -1250)$ chosen in C1Q2.2, choose between $(625, -1375)$ and $(500, -500)$]

Screen Break

[C1Q4.1: If $(500, -500)$ chosen in C1Q3.1, choose between $(1450, -550)$ and $(500, -500)$]

[C1Q4.2: If $(1375, -625)$ chosen in C1Q3.1, choose between $(1300, -700)$ and $(500, -500)$]

[C1Q4.3: If $(500, -500)$ chosen in C1Q3.2, choose between $(1200, -800)$ and $(500, -500)$]

[C1Q4.4: If $(1125, -875)$ chosen in C1Q3.2, choose between $(1050, -950)$ and $(500, -500)$]

[C1Q4.5: If $(500, -500)$ chosen in C1Q3.3, choose between $(950, -1050)$ and $(500, -500)$]

[C1Q4.6: If $(875, -1125)$ chosen in C1Q3.3, choose between $(800, -1200)$ and $(500, -500)$]

[C1Q4.7: If $(500, -500)$ chosen in C1Q3.4, choose between $(700, -1300)$ and $(500, -500)$]

[C1Q4.8: If $(625, -1375)$ chosen in C1Q3.4, choose between $(550, -1450)$ and $(500, -500)$]

Screen Break

[Comparisons 2-6 are identical to Decision Screen 1, with the following exceptions]

[In Comparison 2, the Recipients are: Person B: $35,000 and C: $70,000]

[In Comparison 3, the Recipients are Person C: $70,000 and D: $100,000]

[In Comparison 4, the Recipients are Person C: $70,000 and. E: $170,000]

[In Comparison 5, the Recipients are Person C: $70,000 and F: $250,000]

[In Comparison 6, the Recipients are Person C: $70,000 vs. G: $500,000]

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

Screen Break

[Block 3: Support for Redistribution]

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

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; 2:; 3:; 4: Stay the same; 5; 6; 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; 2; 3; 4; 5:; 6:; 7: Government should not concern itself with income differences]

Screen Break

End of survey

Thank you for your time!

You will be automatically redirected in 5 seconds.

H.2 Treatment Loss x 500K

[All questions are identical to those in Treatment Loss x 70K. The incomes of the Recipients in the comparisons are as follows:]

[In Comparison 2, the Recipients are: Person A: $8,000 vs. G: $500,000]

[In Comparison 2, the Recipients are: Person B: $35,000 vs. G: $500,000]

[In Comparison 3, the Recipients are Person C: $70,000 vs. G: $500,000]

[In Comparison 4, the Recipients are Person D: $100,000 vs. G: $500,000]

[In Comparison 5, the Recipients are Person E: $170,000 vs. G: $500,000]

[In Comparison 6, the Recipients are Person F: $250,000 vs. G: $500,000]

[For half the participants, the order of the Comparisons is reversed]

H.3 Treatment Gain x 70K

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

[Instructions screen]

Instructions

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

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

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

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

Question 2/4: Please choose your preferred alternative

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

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

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

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

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

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

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

[True; False]

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

[True; False]

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

Screen Break

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

[C1Q1: Architect chooses between $(2500, 500)$ and $(2000, 1000)$]

Screen Break

[C1Q2.1: If $(2000, 1000)$ chosen in C1Q1, choose between $(2750, 750)$ and $(2000, 1000)$]

[C1Q2.2: If $(2500, 500)$ chosen in C1Q1, choose between $(2250, 250)$ and $(2000, 1000)$]

Screen Break

[C1Q3.1: If $(2000, 1000)$ chosen in C1Q2.1, choose between $(2875, 875)$ and $(2000, 1000)$]

[C1Q3.2: If $(2750, 750)$ chosen in C1Q2.1, choose between $(2625, 625)$ and $(2000, 1000)$]

[C1Q3.3: If $(2000, 1000)$ chosen in C1Q2.2, choose between $(2375, 375)$ and $(2000, 1000)$]

[C1Q3.4: If $(2250, 250)$ chosen in C1Q2.2, choose between $(2125, 125)$ and $(2000, 1000)$]

Screen Break

[C1Q4.1: If $(2000, 1000)$ chosen in C1Q3.1, choose between $(2950, 950)$ and $(2000, 1000)$]

[C1Q4.2: If $(2875, 875)$ chosen in C1Q3.1, choose between $(2800, 800)$ and $(2000, 1000)$]

[C1Q4.3: If $(2000, 1000)$ chosen in C1Q3.2, choose between $(2700, 700)$ and $(2000, 1000)$]

[C1Q4.4: If $(2625, 625)$ chosen in C1Q3.2, choose between $(2550, 550)$ and $(2000, 1000)$]

[C1Q4.5: If $(2000, 1000)$ chosen in C1Q3.3, choose between $(2450, 450)$ and $(2000, 1000)$]

[C1Q4.6: If $(2375, 375)$ chosen in C1Q3.3, choose between $(2300, 300)$ and $(2000, 1000)$]

[C1Q4.7: If $(2000, 1000)$ chosen in C1Q3.4, choose between $(2200, 200)$ and $(2000, 1000)$]

[C1Q4.8: If $(2125, 125)$ chosen in C1Q3.4, choose between $(2050, 50)$ and $(2000, 1000)$]

[The questions in the other Comparisons are identical to those in Comparison 1]

H.4 Treatment Gain x 500K

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

I Instructions - Study 3

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

I.1 Treatment Real

[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?

Screen Break

[Screen shown if participant does 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.

Screen Break

[Block 1: Background Questions]

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]

Screen 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?

Screen Break

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

End of survey

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

Thank you for your time.

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

Screen Break

[Attention check]

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]

Screen Break

[Block 2: Eliciting Welfare Weights]

[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

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

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:

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.]

Screen Break

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

End of survey

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

You cannot continue with the study.

Thank you for your time.

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

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

End of survey

Thank you for your time.

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

Please click the following link to finish the survey.

Screen Break

[C1Q1: Comparison 1 Question]

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 ○

Screen Break

[All questions hereafter in Comparison 1 look like C1Q1]

[C1Q2.1: If $(500, -500)$ chosen in C1Q1, choose between $(1250, -750)$ and $(500, -500)$]

[C1Q2.2: If $(1000, -1000)$ chosen in C1Q1, choose between $(750, -1250)$ and $(500, -500)$]

Screen Break

[C1Q3.1: If $(500, -500)$ chosen in C1Q2.1, choose between $(1375, -625)$ and $(500, -500)$]

[C1Q3.2: If $(1250, -750)$ chosen in C1Q2.1, choose between $(1125, -875)$ and $(500, -500)$]

[C1Q3.3: If $(500, -500)$ chosen in C1Q2.2, choose between $(875, -1125)$ and $(500, -500)$]

[C1Q3.4: If $(750, -1250)$ chosen in C1Q2.2, choose between $(625, -1375)$ and $(500, -500)$]

Screen Break

[C1Q4.1: If $(500, -500)$ chosen in C1Q3.1, choose between $(1450, -550)$ and $(500, -500)$]

[C1Q4.2: If $(1375, -625)$ chosen in C1Q3.1, choose between $(1300, -700)$ and $(500, -500)$]

[C1Q4.3: If $(500, -500)$ chosen in C1Q3.2, choose between $(1200, -800)$ and $(500, -500)$]

[C1Q4.4: If $(1125, -875)$ chosen in C1Q3.2, choose between $(1050, -950)$ and $(500, -500)$]

[C1Q4.5: If $(500, -500)$ chosen in C1Q3.3, choose between $(950, -1050)$ and $(500, -500)$]

[C1Q4.6: If $(875, -1125)$ chosen in C1Q3.3, choose between $(800, -1200)$ and $(500, -500)$]

[C1Q4.7: If $(500, -500)$ chosen in C1Q3.4, choose between $(700, -1300)$ and $(500, -500)$]

[C1Q4.8: If $(625, -1375)$ chosen in C1Q3.4, choose between $(550, -1450)$ and $(500, -500)$]

Screen Break

[Comparisons 2-6 are identical to Decision Screen 1, with the following exceptions]

[In Comparison 2, the Recipients are: Person B: \$35,000 and C: \$70,000]

[In Comparison 3, the Recipients are Person C: \$70,000 and D: \$100,000]

[In Comparison 4, the Recipients are Person C: \$70,000 and. E: \$170,000]

[In Comparison 5, the Recipients are Person C: \$70,000 and F: \$250,000]

[In Comparison 6, the Recipients are Person C: \$70,000 vs. G: \$500,000]

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

Screen 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]

Screen Break

[**Block 3: Welfarist and Non-Welfarist Concerns**]

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]

Screen 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]

Screen Break

[**Block 4: Knowledge**]

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?

Screen Break

[**Block 4: Mechanisms**]

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.

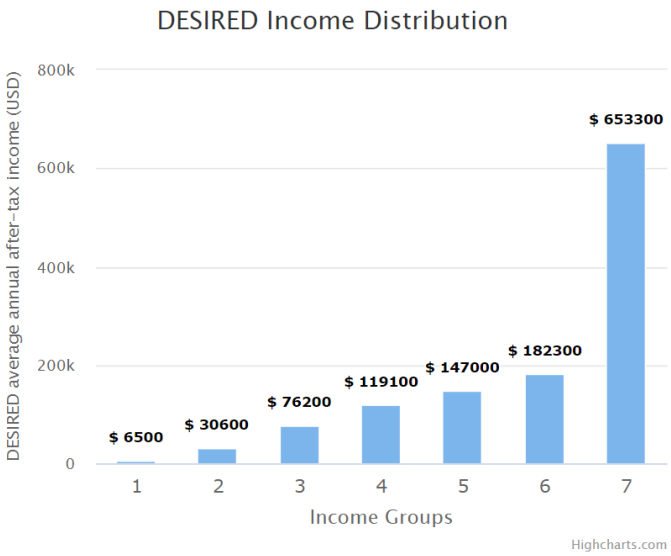
We would like you to indicate what you think the average tax rate for each tax group in the U.S. should be. This can be done as follows. **You can increase or decrease the average tax rates of the first six groups. The average tax rate of group 7 adjusts automatically so that all seven groups together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

Screen 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]

Screen 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.

I.2 Treatment Hypothetical

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

[Block 2: Eliciting Welfare Weights]

[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.]

Screen Break

[C1Q1: 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 Comparisons and questions and identical to those in Treatment Real. Only the first sentence differs between the two treatments]

I.3 Treatment No Self-Interest

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

[**Block 1: Background Questions**]

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

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

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

Screen 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?

Screen Break

[**Block 2: Eliciting Welfare Weights**]

[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:

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**

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

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 instruc-

tions. 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.]

Screen Break

[C1Q1: shown to all participants]

Decision Screen 1/6

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

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

Question 1/4: Please choose your preferred alternative:

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

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

In Comparison 2, the Recipients are Person B: $22,000 to $53,000 and C: $53,000 to $85,000

In Comparison 3, the Recipients are Person C: $53,000 to $85,000 and D: $85,000 to $135,000

In Comparison 4, the Recipients are Person C: $53,000 to $85,000 and E: $135,000 to $210,000

In Comparison 5, the Recipients are Person C: $53,000 to $85,000 and F: $210,000 to $375,000

In Comparison 6, the Recipients are Person C: $53,000 to $85,000 and G: $375,000 and above

[For half the participants, the order of the Comparisons is reversed]

I.4 Treatment Self-Interest

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

[Block 2: Eliciting Welfare Weights]

[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.]

Screen Break

[C1Q1: Comparison 1 Question]

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 Comparisons are identical to those in Treatment No Self-Interest except that in the relevant Comparisons, we replace "Person [A/B/C/D/E/F/G]" with "You." Furthermore, the first sentence in all Comparisons is different.]

References

Bursztyn, L., Handel, B. R., Jimenez, R. & Roth, C. (2023), 'When product markets become collective traps: The case of social media'. NBER Working Paper no. 31771.

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.

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

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

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

Luttmer, E. F. & Samwick, A. A. (2018), 'The welfare cost of perceived policy uncertainty: Evidence from social security', *American Economic Review* **108**(2), 275–307.

Mankiw, N. G., Weinzierl, M. & Yagan, D. (2009), 'Optimal taxation in theory and practice', *Journal of Economic Perspectives* **23**(4), 147–174.

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., 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.

Stantcheva, S. (2021), 'Understanding tax policy: How do people reason?', *The Quarterly Journal of Economics* **136**(4), 2309–2369.

Støstad, M. N. & Cowell, F. (2024), 'Inequality as an externality: Consequences for tax design', *Journal of Public Economics* **235**, 105139.