Wages, taxes, and labor supply elasticities: The role of social preferences*

Janjala Chirakijja[1] and Pinchuan Ong[2]

[1]Monash University, Department of Econometrics and Business Statistics
[2]National University of Singapore, NUS Business School

March 20, 2025

ABSTRACT: We show that social preferences (e.g. warm glow or spite) towards tax-funded government expenditures induce differences between the wage and net-of-tax rate elasticities of labor supply in canonical models. In a large-scale vignette experiment in the US, we find that wage elasticities of labor supply are meaningfully larger than their net-of-tax rate counterparts, consistent with positive social preferences. We show relevance for real labor market decisions by building on an existing meta-analysis of the elasticity of taxable income. Hence, models calibrated using net-of-tax rate elasticities when wage elasticities are more suitable understate the labor supply response of individuals.

JEL: J22, H24, H41.

*Chirakijja: janjala.chirakijja@monash.edu; Ong: ong.pinchuan@nus.edu.sg. We thank Sascha Becker, Marianne Bertrand, Adeline Delavande, Jason Faberman, Lorenz Goette, Jie Gong, Naijia Guo, Jon Guryan, Hilary Hoynes, Anders Humlum, Supreet Kaur, Patrick Kline, John Klopfer, Matthew Notowidigdo, Amanda Pallais, Jessica Pan, Simon Quach, Jesse Rothstein, Benjamin Schoefer, Uta Schoenberg, Dmitry Taubinsky, Basit Zafar, Anson Zhou, James Ziliak, and participants of the NUS Applied Micro SIG Brownbag, NUS Labor Mini-Conference, Chinese University of Hong Kong Lunch Brownbag, University of Hong Kong seminar, Society of Labor Economics 2024 Meeting, Berkeley Labor Lunch, Australasian Public Choice Conference, and Puey Ungphakorn Institute for Economic Research seminar for valuable comments and suggestions. We are grateful to Carina Neisser for making her paper's replication kit publicly available. Justin Goh and Ke Li provided outstanding research assistance. Research reported in this paper received financial support from the MOE AcRF Tier 1 Grant A-0003317-00-00. This study received IRB Approval from the NUS Business School under Protocol SP-23-0103 and was pre-registered as AEARCTR-0011052.

1 Introduction

Economists are generally interested in understanding the responsiveness of labor supply to wages and taxes—the labor supply elasticity. This elasticity plays a key role in business cycle models and for assessing the efficiency costs of labor income taxation, and differences in its magnitude can give rise to different conclusions (Keane 2011).

When estimating the labor supply elasticity, economists conventionally do not distinguish between responses to wages and taxes. Despite a nominal distinction—a response to taxes is more accurately called an elasticity with respect to the net-of-tax rate (e.g. see Saez et al. 2012)—virtually all models of labor supply used for estimation assume that people care about wages and income taxes only through their effects on disposable income. Hence, the conventional thought is that an increase in wages and a decrease in the tax rate have equivalent effects on disposable income, which imply equivalent effects on labor supply, and in practice, most empirical estimates of the labor supply elasticity are based on responses to after-tax wages (see e.g. Eissa and Hoynes 2004; Ziliak and Kniesner 2005; Blau and Kahn 2007; Keane and Wasi 2016; Blundell et al. 2016; Unrath 2020). In general, the only studies that do not assume this interchangeability between taxes and wages (or, more generally, prices) are those that study differential salience between the two (Chetty et al. 2009; Finkelstein 2009; Blumkin et al. 2012; Taubinsky and Rees-Jones 2018; Kroft et al. 2024).

This paper identifies another reason (besides salience) to distinguish between estimates of the wage and the net-of-tax rate elasticities of labor supply. A large literature points to the existence of other-regarding or social preferences (Andreoni 1989, 1990; Fehr and Schmidt 1999; Charness and Rabin 2002) in many settings. Applied to income taxation by the government, people might like or dislike the fact that their tax dollars eventually go towards (or are wasted in the course of providing) public goods expenditure. In addition, unlike wage changes which are often viewed as being specific to the individual, tax rate changes almost always apply to many people at once (e.g. all taxpayers in the same taxable income band). Hence, tax changes potentially affect the aggregate provision of public goods which individuals may value. Such government-related social preferences would lead individuals to derive utility or disutility from income taxation over and above what would be explained by its impact on their disposable income.

We show that incorporating government-related social preferences in a labor supply model drives a wedge between the responses of labor supply to wages and the net-of-tax rate. Intuitively, if people derive satisfaction from their or others' contributions to public good provision induced by an income tax, the tax increase no longer hurts the individual as

much as an equivalent decrease in wages—while disposable income is reduced, the individual gets some consolation from the fact that taxes eventually serves some good. Taken to the extreme, if the individual derives as much utility from tax-funded government expenditure as consumption from disposable income, the response to tax changes could be very small. On the other hand, a wage increase is good for both the government budget and for own consumption. This implies that if people derive satisfaction from their tax dollars, the wage elasticity of labor supply should be larger than the net-of-tax rate elasticity of labor supply.

To estimate the size of this wage-tax elasticity wedge, we field a vignette experiment on a large sample of US residents recruited through an online panel.[1] A vignette experiment (hypothetical choice experiment) is the preferred methodology for our purpose since it allows us to obtain causal estimates of the two different elasticities (wage or net-of-tax rate) for the same population. Conventional methods using observational data are unlikely to achieve this.[2] Our estimates of the wedge have high internal validity since we control the randomization process, and we use appropriate wording in our questions and instructions to ensure that participants hold market frictions and other factors constant when making their choices. In particular, information about wage and tax rates are designed to be equally salient to respondents, hence removing salience as the reason for our results. Our vignette experiment is similar in spirit to Kosar et al. (2019), who randomly vary wages and work hours in their scenarios to study heterogeneity in labor supply elasticity estimates across demographic groups.[3] To fit our purpose, we additionally vary a tax rate across scenarios.

We find that respondents, on average, exhibit positive social preferences towards tax-funded government expenditures. To estimate the wage-tax elasticity wedge, we first estimate the structural parameters of a logit choice model and then simulate the elasticities of interest. Consistent with positive social preferences, we find that estimated wage elasticities are meaningfully larger than net-of-tax rate elasticities in our vignette experiment. In our

[1]For brevity, we use the term "wage-tax elasticity wedge" to denote the difference between the wage elasticity of labor supply and the net-of-tax rate elasticity of labor supply (the former minus the latter).

[2]For example, suppose we used observational data and two instruments—one for wages and one for tax rates—to estimate the two types of labor supply elasticities. This has the advantage of being immediately relevant to real market decisions. However, even if we obtained both elasticity estimates in the same sample, there is no guarantee that the compliers to the two different instruments (Imbens and Angrist 1994; Angrist et al. 1996) are the same people. This would result in a biased estimate of the wedge.

[3]Vignette experiments (sometimes known as conjoint analysis) have been used in labor economics to estimate preferences for specific nonpecuniary job amenities or disamenities (Eriksson and Kristensen 2014; Mas and Pallais 2017; Wiswall and Zafar 2018; Mas and Pallais 2019; Kosar et al. 2019; Folke and Rickne 2022; Maestas et al. 2023; Burbano et al. forthcoming), including distaste for work via the inclusion of work hours in the list of vignette disamenities (Wiswall and Zafar 2018; Mas and Pallais 2019; Kosar et al. 2019; Maestas et al. 2023). However, Kosar et al. (2019) and Mas and Pallais (2019) are the only prior studies to report labor supply elasticities—the relevant parameter used in many applications—after estimating the distaste for work.

heterogeneity analysis, the wage-tax elasticity wedge is larger if the respondent has a better opinion of the government and its programs, consistent with the wedge being larger if the respondent has stronger social preferences. We also find some evidence that government-related social preferences can be usage-specific: the wage-tax elasticity wedge is larger if taxes are earmarked for specific programs that the respondent likes more.

Since our experimental results are based on stated preference, a natural concern is how externally valid they are for choices that workers actually have to make in the labor market. To address this, we build on a recent meta-analysis of the elasticity of taxable income (ETI) with respect to the net-of-tax rate by Neisser (2021). Because it examines taxable income rather than measures of labor supply (e.g. work hours), the ETI generalizes the standard labor supply elasticity by additionally capturing other behavioral responses (e.g. compensation timing, tax avoidance) that are relevant for analysis of income taxation (Feldstein 1995; Saez et al. 2012).[4] Important for our purposes, government-related social preferences should also be reflected in the ETI—a better opinion of the government should be correlated with a lower ETI. Using Neisser's replication kit, we examine the relationship between the ETI and variables from the World Values Survey (WVS) and the European Values Survey (EVS). We find that the ETI is negatively correlated with variables that capture institutional trust in these surveys, even after controlling for a large number of study-related covariates used in Neisser's study. This suggests that government-related social preferences matter for the labor supply response to taxation, which is a policy-relevant finding even ignoring our previous discussions about the wage-tax elasticity wedge.

Our paper contributes to a large literature on the labor supply elasticity (see Keane 2011, 2022 for reviews). Previous work has estimated this elasticity by using variation in wages (MaCurdy 1981; Altonji 1986; Camerer et al. 1997; Oettinger 1999; Pistaferri 2003; Ziliak and Kniesner 2005; Farber 2005; Blau and Kahn 2007; Fehr and Goette 2007; Stafford 2015; Giné et al. 2017; Chen et al. 2019) or taxes (Eissa 1995; Blundell et al. 1998; Bianchi et al. 2001; Eissa and Hoynes 2004; Chetty et al. 2011; Gelber 2014; Blundell et al. 2016; Sigurdsson 2019; Stefánsson 2019; Unrath 2020; Martinez et al. 2021; Elder et al. 2023; Sigaard 2023), treating the response to the two as equivalent. However, a small but growing research suggests that people may respond differently to wage versus income tax changes for reasons other than salience, although the evidence is still mixed. Using real-effort experiments on smaller samples, Kessler and Norton (2016) find more negative worker-effort responses to the introduction of a tax compared to an equivalent wage cut, whereas Mori et al. (2022) find

[4]ETI estimates are always net-of-tax rate estimates and hence implicitly capture social preferences. That said, "behavioral response" in this literature refers to the fact that the outcome (taxable income) is more general than just labor supply. We are unaware of any study that discusses the social preferences reasons for variation in the ETI.

no wage-tax differences when they increase the tax rate from a non-zero base, and Esslinger et al. (2024) find no differences for tax cuts and some differences for tax increases. Djanali and Sheehan-Connor (2012) and Rick et al. (2018) find larger responses to wage (i.e. in line with our results). While we do not use a real-effort experiment, our approach is designed to better reflect tax variation outside of a laboratory—we use a more nationally representative sample, our vignettes use tax variation that are closer to policy-relevant magnitudes, and our tax variation reflects how tax changes often affect many people in the same tax band. A separate body of work have argued and documented using administrative data that tax-benefit linkages shift the incidence of payroll taxes towards employees, implying that workers perceive these benefits when they make their labor supply decisions (Summers 1989; Gruber 1997; Bozio et al. 2019). We generalize beyond a linkage that benefits the self to social preferences, and focus on the labor supply decisions of workers directly. More generally, we provide a theory—most similar to Djanali and Sheehan-Connor (2012)—that could be used to think about the results of the studies discussed here.

A related literature investigates whether estimates of the intertemporal elasticity of labor supply (the Frisch elasticity) based on microdata can be used to calibrate the (intertemporal) labor supply response to business cycle fluctuations. Previous work has argued that estimates based on life-cycle models are underestimated due to credit constraints (Domeij and Floden 2006), precautionary savings (Low 2005), endogenous human capital accumulation (Imai and Keane 2004; Keane and Wasi 2016), optimization frictions (Chetty 2012), and the decision not to model the participation margin (Chang and Kim 2006; Rogerson and Wallenius 2009; Keane and Wasi 2016). Recent Frisch elasticity estimates based on the response to a tax-free year of income are designed to be robust to all the above (Martinez et al. 2021; Sigurdsson 2019; Stefánsson 2019).[5] These estimates range between 0.025 in Switzerland and 0.4 in Iceland, and are too small to explain the variation in work hours over the business cycle. Our findings partially explains the discrepancy—labor supply responds to the wage signal of the business cycle, and hence should be larger than estimates based on tax variation. Similar considerations apply to estimates based on life-cycle models that exploit exogenous variation in taxes for identification of Frisch elasticities (e.g. Blundell et al. 2016). This is particularly relevant when we notice that quasi-experimental estimates tend to be based on tax variation, possibly because national-level tax changes are more easily observed by the econometrician than the wage offers of individual employers.

Our paper draws insights from a large literature on how social preferences interact with

[5]Specifically, the studies argue that the relevant frictions in their settings would also apply to business cycle models, and hence macroeconomic models *should* be using the "frictional" Frisch. Furthermore, since they examine the response of a whole population to the tax rate change instead of following individuals over their life cycles, they avoid issues of precautionary savings and endogenous human capital accumulation.

taxation and public goods provision. This literature has focused mainly on intrinsic motivations for tax compliance and tax morale (Frey 1992; Konrad and Qari 2012; Lamberton et al. 2018; Nathan et al. 2021; Doerrenberg and Peichl 2022; Cingl et al. 2023; Besley et al. 2023), with an active sub-literature emphasizing the psychological link between taxation and public goods provision (Cowell and Gordon 1988; Alm et al. 1993; Hall and Preston 2000; Cullen et al. 2021; Giaccobasso et al. 2022; Falsetta et al. 2023).[6] More broadly, prior studies have studied social preferences in the context of charitable giving (DellaVigna et al. 2012; Andreoni and Payne 2013; Exley 2016; Ottoni-Wilhelm et al. 2017; Carpenter 2021), volunteering (Freeman 1997; Bauer et al. 2013; Lilley and Slonim 2014), gift exchange and worker effort (Krueger and Mas 2004; Gneezy and List 2006; Mas 2006; Kube et al. 2012, 2013; DellaVigna et al. 2022), redistribution (Alesina and Angeletos 2005; Alesina and Giuliano 2011; Luttmer and Singhal 2011; Huber and Stanig 2011; Durante et al. 2014; Kuziemko et al. 2015; Karadja et al. 2017; Stantcheva 2021; Almås et al. 2022; Hvidberg et al. 2023), and optimal taxation (Saez 2004; Diamond 2006). We contribute to the broader social preferences literature by demonstrating its importance for non-volunteering labor supply. Additionally, this study contributes to our understanding of intrinsic motivations for paying taxes—our respondents derive utility from taxes paid, and it increases if the taxes fund programs that the respondents likes. This is potentially important for strategies to study tax morale since labor supply is more easily observed in data than tax compliance.

The paper proceeds as follows. Section 2 formalizes the intuition for the wage-tax elasticity wedge in a model. Section 3 describes our survey, vignette experiment, and estimation strategy. Section 4 presents results of our vignette experiment. Section 5 presents correlations between the elasticity of taxable income and proxies for government-related social preferences. Section 6 concludes.

2 Model

In this section, we incorporate government-related social preferences in the canonical dynamic model of labor supply, and derive implications for the wage and net-of-tax rate elasticities of labor supply. We use the elasticities from this section later when we estimate the wage-tax elasticity wedge in Section 4.

In any time period, an individual i with initial assets level a_i chooses consumption c_i,

[6] See Andreoni et al. (1998), Luttmer and Singhal (2014), and Alm (2019) for reviews from the perspective of tax compliance.

next period assets a_i', and labor supply h_i to solve the utility maximization problem[7]

$$V\left(a_i, w_i, \tau, N_i\right) = \max_{c_i, a_i', h_i} U\left(c_i, h_i, g_i\right) + \beta E\left[V\left(a_i', w_i', \tau', N_i'\right)\right] \tag{1}$$

$$\text{s.t. } c_i + \frac{1}{1+r} a_i' = a_i + \left(1-\tau\right) w_i h_i + N_i, \tag{2}$$

$$g_i = \tau w_i h_i, \tag{3}$$

where w_i is the person's wage, τ is the tax rate on earnings from work $w_i h_i$, N_i is non-labor income, β is an impatience parameter, r is the interest rate, $V\left(\cdot\right)$ is the indirect utility function, and $E\left[\cdot\right]$ is the expectations operator. Equation (2) is a usual budget constraint which specifies that consumption c_i and savings $\frac{1}{1+r} a_i' - a_i$ equals post-tax earnings and non-labor income.

We make only one change compared to the usual labor supply model: per-period utility in Equation (1) also depends on the amount of government expenditure from the individual's taxes paid g_i (tax-funded government expenditure), equivalent to the tax rate applied to their total earnings from work by Equation (3). In our main analysis, we specialize this utility function to

$$U\left(c_i, h_i, g_i\right) = u\left(c_i, h_i\right) + v\left(g_i\right) \tag{4}$$

so that we have an easily interpretable social preferences term $v\left(g_i\right)$. Similar to DellaVigna et al. (2012), we use a general $v\left(g_i\right)$ that can encompass both pure altruism (people value the public goods being provided) or warm glow (people like contributing to nation building, regardless of the value of the public goods provided), and we do not take a stand on which is more important. We also allow for both positive and negative social preferences. We expect individuals to value tax-funded public goods to some degree, and there might be diminishing marginal returns to public goods provision; this translates to $v\left(g_i\right)$ having positive and negative first and second derivatives, respectively. However, in principle, $v\left(g_i\right)$ might have a negative first derivative, if the individual thinks that public goods provision is net-detrimental to society (e.g. anarchists might dislike tax-funded institutions) or himself (e.g. criminals might dislike tax-funded police activity).

Note that we model the tax rate τ as a common tax rate that does not depend on i. Since tax rate changes often affect broad groups of people in similar circumstances, individuals may also get social preferences utility from the increased (or decreased) contributions of others, in addition to their own. While our vignette experiment was not designed to vary the tax contribution of others separately from the individual's own contribution, we show in

[7] We use the prime symbol to denote the next period for notational simplicity.

Appendix B.1 that the model in this section can be derived from a more complex model with utility that depends separately on own and others' contributions, as well as total government expenditure. Hence, $v(g_i)$ can be interpreted to approximately capture social preferences from others' contribution as well.

In Appendix B.2, we derive the form for the three types of labor supply elasticities frequently used by economists: the uncompensated Marshallian elasticity, the Hicksian elasticity which compensates for income effects, and the Frisch elasticity which holds the marginal utility of wealth constant. In general, for all three elasticity types, the wage elasticity differs from the net-of-tax rate elasticity. For example, denoting partial derivatives using subscripts and dropping utility's dependence on its arguments and i, the Frisch wage elasticity is

$$\epsilon_w^F = \frac{-w_i(1-\tau)U_{cc}\left[U_c + \frac{\tau}{1-\tau}(U_g + U_{gg}g)\right]}{h_i\left(U_{cc}U_{hh} - U_{ch}^2 + \tau^2 w_i^2 U_{cc}U_{GG}\right)} \tag{5}$$

while the Frisch net-of-tax rate elasticity is

$$\epsilon_{1-\tau}^F = \frac{-w_i(1-\tau)U_{cc}\left[U_c - (U_G + U_{gg}g)\right]}{h_i\left(U_{cc}U_{hh} - U_{ch}^2 + \tau^2 w_i^2 U_{cc}U_{gg}\right)}. \tag{6}$$

The key difference between the two elasticities is the sign on the $U_g + U_{gg}g$ term. This term captures the intuition for the wage-tax wedge. Typical concavity assumptions on $U(\cdot)$ ensure that $\frac{-w_i(1-\tau)U_{cc}}{h_i\left(U_{cc}U_{hh}-U_{ch}^2+\tau^2 w_i^2 U_{cc}U_{gg}\right)}$ and U_c are positive, with their product capturing the usual intuition that increases in disposable earnings encourage substitution towards working. For the Frisch wage elasticity, $U_g + U_{gg}g$ reforces this substitution effect, assuming that taxes were not so high that diminishing marginal utility starts to kick in. For the Frisch net-of-tax rate elasticity, $U_g + U_{gg}g$ acts in opposite direction to the substitution effect, because an increase in tax rates funds more public goods even as it reduces disposable earnings.

The immediate implication of the above is that social preferences cause the Frisch wage elasticity to be larger than the Frisch net-of-tax rate elasticity, with the wedge being larger if social preferences is stronger. This is true in general for all three types of elasticities. We next turn to our vignette survey experiment to estimate the size of the wedge.

3 Survey, vignette experiment, and estimation strategy

Our empirical goal is to estimate how individuals' labor supply decisions react to wage changes compared to tax changes. Since observed labor supply choices are influenced by various other determinants, it is difficult to identify the labor supply elasticities using observational data without imposing some additional assumptions. To overcome this challenge,

we designed a vignette experiment that allows us to obtain causal estimates for the two types of elasticities (wages and net-of-tax rates) that are internally consistent. Our experimental design was inspired by Kosar et al. (2019) which estimated labor supply elasticities using a vignette method in order to investigate heterogeneity across demographic groups. The key advantages of this approach are that the variations are all exogenous, and we do not need to make assumptions about other factors that affect individuals' work hours. Throughout the survey, we employed several approaches to ensure that respondents understood and paid attention to the questions we asked—Appendix C.1 discusses the details. We provide webpage links to a live version of our online survey and the full questionnaire in Appendix D.

3.1 Survey structure and sample

We conducted a large-scale vignette experiment embedded in an online survey fielded between February and June 2023. Our survey targeted US residents aged 25–64 who were working. We restricted participation to those working for wage income and excluded self-employed workers—we asked respondents to think about scenarios where they had to find a new job with similar conditions, and self-employed individuals might have difficulty relating to this. We recruited participants via Bilendi/Respondi, a survey company commonly used in social science research (see e.g. Alesina et al. 2021; Stantcheva 2021; Jäger et al. 2023).[8] Our survey was approximately 15 to 30 minutes long, with a median time for completion of 20.1 minutes. We required that participants access our survey on their laptop or tablet devices only (as opposed to mobile phones), due to the size of our vignette screens.

Our survey was structured as follows:

1. Demographic and employment questions: We started by collecting information on key demographic characteristics. We then asked respondents about their employment status, earnings, and household wealth, including work hours, wage/salary income of self and spouse, assets and liabilities, usual monthly consumption, and financial resources available if unemployed. The survey instrument was programmed to use the information collected in this section to compute each respondent's hourly wage and simulate their income tax rate, needed for our vignette experiment.[9] Appendix C.3

[8] The company sends out survey links via email to respondents who participate in opinion surveys in exchange for money or reward points. Stantcheva (2023) describes the recruitment strategies generally employed by survey companies and how their respondent pools compare with the broader population. Of note, participants were informed that their responses would be subject to statistical checks, and that low-quality responses might be dropped from our study.

[9] Respondents were asked to verify the hourly wage and amend their reports of income or work hours if the hourly wage did not make sense. Simulated tax rates were not similarly verified to avoid emphasizing taxes early in the survey. We did not ask respondents directly about their taxes paid for the same reason.

provides details about our tax simulation procedure.

2. Vignette experiment: Respondents first clicked through an animated slideshow that walked them through the instructions for the vignettes. This set of ten non-skippable slides explained the hypothetical scenarios, the assumptions the respondents should be making and the choices they would be asked to make. Respondents then proceeded to answer the questions in our vignette experiment, described in Section 3.2.

3. Views on government and tax questions: After completing the vignette experiment, we elicited opinions on the federal government and tax system. These included questions about the perceived fairness of and satisfaction with the federal tax system, and attitudes towards the role and capacity of the government, particularly in how it spends taxpayer's money. In addition, respondents were asked the extent to which they liked or disliked that their tax money is used to fund different spending categories such as national defense, international affairs, transportation, health insurance programs, and social security. The survey closed with miscellaneous questions, including an open-ended feedback question on the survey.

7,122 respondents completed our survey. To arrive at our main analysis sample, we drop individuals whose work hours, consumption, or tax amounts are below the 5th or above the 95th percentiles, to avoid the influence of outliers on our simulated elasticities.[10] We also exclude a small number of respondents who skipped our compulsory instructions slide show—this could be achieved if the respondent knew how to manipulate Javascript in a browser console—and respondents whose consumption information in the vignettes cannot be imputed (e.g., if they always reported the maximum or minimum values on the consumption scale).

Table 1 shows the characteristics of our final sample of 5,440 individuals and those of the US working population. The population statistics come from the Current Population Survey. By construction, our sample is representative along the targeted dimensions of age, gender, and income. There is also broad similarity along some non-targeted dimensions such as marital status and race, although the sample is slightly more educated, less Hispanic, and more Democrat.[11] Appendix C.2 provides more details about our sample.

[10] For example, if the the utility function in Equation (1) contained $\log c$, the elasticities in Equation (5) and (6) would contain $1/c$ terms. Observations with very small consumption values would then have a big influence on the elasticities that we simulate.

[11] Stantcheva (2022, Table 1) also recruited a sample of respondents in the labor force via Bilendi/Respondi that was slightly more educated, less Hispanic, and more Democrat than the US population. Hence, this might be a common feature of the online sample. We check robustness to sample non-representativeness in Section 4.2.

3.2 Vignette experimental design

The vignette experiment showed respondents ten pairs of hypothetical scenarios. In all scenarios, respondents were asked to assume that everything about them was the same as their situation at the time of the survey except that they had to leave their job and find a new one. The scenario description read:

> "You have to leave your current job and find a new one. You have received a job offer that will pay you $*[wage]* per hour and require that you work *[hour]* hours per week. You will have to pay the federal income tax on the amount you earn from this job at the rate of *[tax rate]*%."

The scenarios randomized the three bracketed variables in the description above around the respondents' actual values. Wages and tax rates were drawn uniformly from the 21 and 11 even numbers closest to the actual wage and tax rate, respectively, and work hours were drawn uniformly from the 9 multiples of five closest to the actual work hours. We constrained the wage and hours grids to be positive, and the tax rate grid to be nonnegative.[12] In each pair of scenarios, respondents were asked if they would take up the job in each scenario and which of the two scenarios they preferred.

We instructed respondents to assume the following as they thought about the hypothetical scenarios:

- Jobs and family members' situations were otherwise identical to actual situations (except for different hourly wage and work hours specified in the scenarios);
- The income tax rate in the scenarios was the federal income tax, and applied to people "with similar income and employment situation to you";[13]
- The federal government balanced its budget; and
- If they chose not to take the job, they would have access to the same financial resources that are available if they were to quit their actual job, and the same chance of finding a new job.

[12] For example, a respondent earning 31 dollars per hour at a 40 hour-per-week job and facing a tax rate of 8 percent would see wages drawn from 12, 14, ..., 50, and 52, work hours drawn from 20, 25, 30, 35, 40, 45, 50, 55, and 60, and tax rates drawn from 0, 2, ..., 20 (these being the eleven even numbers closest to the actual tax rate while still being nonnegative).

[13] We worded the instructions this way for realism—tax rates often apply to broad groups of people with similar income and employment situations, but yet not everyone in the country pays the same tax rate. Note this still allows for pure altruism to be at play (in addition to warm glow), since a higher tax rate might affect sufficiently many people such that the value of public goods provided increases.

Appendix Figure A1 shows the basic layout of our scenario pairs. To make our vignettes as easy to understand and respond to as possible, we computed and explained the workings for all relevant details—specifically the monthly pre-tax earnings, take-home income, and tax amounts paid. We also required participants to answer the questions step-by-step, starting with the choice of whether to take up the job in each of the two scenarios before selecting their preferred scenario. Additionally, we repeated the instructions at the top of the screen for respondents' reference.

Pairs 1 to 4 followed the basic layout. Pairs 5 to 8 involved an additional step—delayed to minimize cognitive load—by also collecting information on consumption, which we need for estimation.[14] Specifically, we asked respondents how much their household would spend in total each month if they took up the job in each scenario. Pairs 9 and 10 added additional features, namely an earmarked tax for a specific program, that allow us to investigate whether the size of the wage-tax elasticity wedge varies with whether the taxes fund specific programs that respondents like. We discuss the design of pairs 9 and 10 together with the analysis in Section 4.3. The vignette part of our survey is necessarily more complex than other sections; to help respondents along and to maintain attention, we showed an interactive "guided website tour" every time we introduced a change in the layout (pairs 1, 5, and 9).

We built in two attention checks in our survey. The first doubles as a comprehension check, and asked participants a simple question about the instruction of the vignette following the instructional slideshow mentioned in Section 3.1 point 2. Specifically, respondents were asked whether the income tax rate in the hypothetical scenarios referred to federal, state or local income tax rate (the correct answer being federal). Second, pair 7 of our vignette experiment presented a strictly dominating relationship, with both scenarios showing the same tax rate and work hours but different hourly wages. We expect attentive participants to prefer the scenario with the higher wage. About 7 and 8 percents of our sample failed the two checks, respectively, a rate comparable with Mas and Pallais (2017) and the literature on inattention in online surveys (see e.g. Peer et al. 2022). (Combined, 13% of respondents failed at least one of the two attention checks).

3.3 Consumption imputation

Estimating the model in Section 2 requires consumption information, which we collected in only pairs 5 to 8 by design. To use information in all pairs, we impute consumption for all scenarios. We first assess if respondents are correctly basing their consumption decisions on take-home earnings rather than pre-tax earnings in pairs 5 to 8. Appendix Table A1 reports

[14] Appendix Figure A2 shows the layout.

that, in a horse-race regression of log consumption on log take-home earnings and log pre-tax earnings, the effect of "more earnings" loads on take-home earnings.[15] We therefore predict log consumption for all scenarios using log take-home earnings at the individual-level.[16] That is, the variation in consumption comes from each respondent's take-home earnings (which varies across scenarios) and a respondent-specific slope based on eight data points collected in pairs 5 to 8.

3.4 Main empirical strategy

Using the vignette data, we estimate the wage and net-of-tax rate elasticities for the three types that are commonly used in the literature (Marshallian, Hicksian, and Frisch). We start by estimating the structural parameters of a choice model using the data. These structural parameters are then used to simulate the elasticities of interest.

In our main analysis, each respondent i saw eight pairs t of scenarios j (pairs 1 to 8 of our vignettes). The scenarios randomly varied three parameters (wage, tax rate, and work hours), which map to consumption c_{itj} (described in Section 3.3), work hours h_{itj}, and tax-funded government expenditure g_{itj} (the last two were presented to the respondent directly). To estimate the structural utility parameters using our vignette experiment, we parameterize the utility function in Equation (4) with

$$\begin{aligned} U\left(c_{itj}, h_{itj}, g_{itj}\right) = {} & \beta_c \log c_{itj} + \beta_h \log\left(\bar{L} - h_{itj}\right) + \beta_{cc}\left(\log c_{itj}\right)^2 \\ & + \beta_{hh}\left(\log\left(\bar{L} - h_{itj}\right)\right)^2 + \beta_{ch} \log c_{itj} \log\left(\bar{L} - h_{itj}\right) \\ & + \beta_g g_{itj} + \beta_{gg} g_{itj}^2 + \beta_{ggg} g_{itj}^3 + \xi_{itj}, \end{aligned} \tag{7}$$

where the β's are coefficients to be estimated, $\bar{L}$ is the maximum time available so that $\bar{L} - h_{itj}$ is the amount of leisure, and ξ_{itj} is a variable that captures all other factors that might affect utility, including possible model misspecification.

The first two lines of Equation (7) model preferences over consumption and leisure with a translog specification that has been used in previous papers (e.g. Hotz et al. 1988; Shaw 1989; Ziliak and Kniesner 2005; Elder et al. 2023). The third line models social preferences as a flexible function of tax-funded government expenditure. Importantly, the polynomial specification for social preferences imposes no restriction on whether respondents like or

[15] Our estimate implies that a 1 percent increase in take-home income is associated with a 0.37 percentage increase in consumption. This elasticity is of a similar magnitude to the elasticity of consumption with respect to permanent income shocks estimated by Baker (2018).

[16] We control for pair and scenario order (whether the scenario is the first or second) fixed effects, and drop observations in which consumption is at the minimum or maximum of the range slider used to collect consumption information.

dislike (in the manner of Levine 1998) tax-funded government expenditure or the degree of diminishing marginal utility (Carpenter 2021).

We model ξ_{itj} as independent and identically distributed Type I extreme value conditional on the respondent-pair, and assume that choices that the respondent made—to work or not for each scenario, and which scenario she preferred—are those which gives the highest utility level. This allows us to estimate the coefficients in Equation (7) using standard fixed-effects logit regressions (McFadden 1974). We additionally control for the order of the scenario (first of the pair shown versus second). Equation (1) includes the expected indirect utility of the next period; we account for this by controlling for $\log\left(assets_i + spouseincome_i + savings_{itj}\right)$ and its square, where $assets_i$ and $spouseincome_i$ are respectively the wealth and spouse's income in dollars of the individual, and $savings_{itj}$ is the after-tax disposable income less consumption for the scenario. We also control for whether the scenario's work hours was below 35 hours per week, following previous papers which have included a part-time fixed cost in their utility functions (Van Soest 1995; Euwals and Van Soest 1999; Elder et al. 2023).[17] Standard errors are clustered at the respondent level.

After estimating the model, we simulate the elasticities and their differences using Equations (B19), (B20), (B31), (B32), (5), and (6). These elasticities are computed using the coefficients from the model (the β's), the respondents' actual (non-vignette) reports of the non-tax variables, and the simulated taxes paid described in Appendix C.3. Standard errors are computed using the delta method.

To investigate heterogeneity with respect to an opinion of the government K_i, we elaborate Equation (7) by allowing the coefficients to depend on K_i:

$$\beta_l = \alpha_{0l} + \alpha_{1l} K_i, \quad l \in \{c, h, cc, hh, ch, g, gg, ggg\} \tag{8}$$

where l is an index of a coefficient in Equation (7). We estimate the model as before, and then simulate the difference between high and low opinions of the government as the average partial effect of K_i on the wage-tax elasticity wedge.

[17] These papers usually also include a fixed cost of working and model coefficient heterogeneity in demographics. The former is not relevant in our analysis because we only showed scenarios with positive work hours. We show robustness to modeling demographic heterogeneity in Section 4.2.

4 Results

4.1 Main vignette experiment

We start by investigating whether respondents indeed exhibit government-related social preferences using a reduced form specification. Our vignette experiment was not designed to show evidence on this in a straightforward manner, but we can still examine social preferences by controlling for the disposable income and work hours of the scenario. Conceptually, holding constant disposable income and work hours leaves us with variation that comes from higher wages that are immediately channeled into taxes, our regressor of interest. A preference for scenarios with higher (lower) taxes would then indicate that our respondents exhibit positive (negative) social preferences towards tax-funded government spending.

Table 2 reports estimates based on regressions of the choices that respondents made against taxes paid in the vignette experiments. An observation is a respondent-pair-scenario, and all columns include individual, pair, and order fixed effects. The first three columns present results for the choice of whether to work or not, and the last three columns present results for the choice of whether the respondent preferred the scenario over the other in the pair. Columns 1 and 4 control for disposable income and work hours flexibly using a a fifth-order bivariate polynomial in the two variables. Columns 2 and 5 control nonparametrically for disposable income and work hours by including fixed effects for cells constructed by interacting centiles of disposable income with centiles of work hours. Columns 3 and 6 further interact these cells with linear slopes in disposable income and work hours, thus protecting against possible within-cell correlations between taxes paid and these two variables.

Controlling for disposable income and work hours, all estimates are positive and significant—respondents were more willing to work and preferred scenarios with higher tax amounts. We interpret this result as evidence that people do derive positive utility from government expenditure. In Figure 1, we show binscatter plots corresponding to Columns 3 and 6 of the table. The relationship appears relatively linear and is not driven by outliers.

We next turn to our main structural estimates of utility parameters in order to investigate the wage-tax elasticity wedge. Table 3 reports estimates of the utility parameters estimated via logit. The first two columns report estimates based on the choice of whether to work in the scenario, and the next two columns report estimates based on the choice of the preferred scenario. Odd columns report estimates ignoring social preferences, and even columns report estimates with a third-degree polynomial in tax-funded government expenditure. The social preferences parameters are quite precisely estimated.

At the bottom of the table, we report the average marginal utilities of consumption and tax-funded government expenditure, and their ratio. Consistent with the reduced

form results, respondents exhibit positive government-related social preferences. The ratio $E(U_g)/E(U_c)$ is smaller than one, implying that respondents still value own consumption more than government expenditure on average.[18]

In the last row, we report $E(U_{gg}g)$, the curvature of social preferences multiplied by the amount of tax-funded government expenditure. Its negative value implies that the social preferences function is concave—the marginal utility of tax-funded government expenditure diminishes as tax payments increase. Diminishing marginal utility is more consistent with warm glow (satisfaction from contributing) than pure altruism (satisfaction from the overall provision of public goods provided), since changes in the individual's contribution should have a negligible impact on overall provision (DellaVigna et al. 2012). By Equations (5) and (6), concavity also reduces the impact of positive social preferences on the wage-tax elasticity wedge. That said, the magnitude of $E(U_g + U_{gg}g)$ is still net positive.

Table 4 reports estimates of the wage elasticities, net-of-tax rate elasticities, and the differences between the two. The columns correspond to those from Table 3, with each cell estimated using the parameters of the model and the actual (or tax-simulated) values of the respondents (see Section 3.4 for details). The three panels each show estimates for a type of elasticity commonly used in the literature, respectively, the Marshallian, Hicksian, and Frisch elasticities of labor supply.

Columns 1 and 3 show the estimates of the elasticities without incorporating social preferences, as assumed by all previous models of labor supply that we are aware of. Without social preferences, the differences between the wage and net-of-tax rate elasticities are necessarily zero. While not the main focus of our paper, it is useful to discuss the magnitude of the elasticity estimates in these two columns. The estimated Marshallian and Hicksian elasticities are within range of what previous papers have estimated (see e.g. Keane 2011; Bargain and Peichl 2016), and the estimated Frisch elasticities are at the high end of what others have estimated in real labor market situations.[19] The latter points to the nature of our estimates. The labor supply elasticities in this paper are better interpreted as frictionless elasticities, since many market frictions—including salience, search frictions (in columns 3 and 4 of the table), and constraints on work hours—should not be in play as our respondents

[18] We can compute analogous ratios for papers on charity giving based on their parameter estimates and evaluated at average values. This yields 0.25 for DellaVigna et al. (2012) and 0.90 for Ottoni-Wilhelm et al. (2017).

[19] The largest estimates of Frisch wage elasticities that we are aware of are 1.2 for bicycle messengers (Fehr and Goette 2007), 1.3 for fishermen (Stafford 2015; Giné et al. 2017), and 1.9 for Uber drivers (Chen et al. 2019). These estimates are based on shorter decision-making time horizons and, as the authors emphasize, are for workers who are more likely to be able to choose their labor supply decisions freely. Additionally, Mui and Schoefer (forthcoming) use representative surveys in the US and Germany to elicit wage changes that would induce a person to work (or stop working), and find that for small wage changes around the respondents' current wages, the implied Frisch elasticities exceed 3.

made their choices.[20] The larger Frisch elasticity estimate in column 3 compared to column 1 also points to the role of frictions. Our experiment specified that if the respondent chose not to take a job, her available resources and job-finding probability would be the same as if she were to quit her actual job. This introduces a job search friction that would affect the choice of whether to work, but not the choice of which scenario she preferred.

We now turn to our key results shown in columns 2 and 4 of the table. With social preferences, the wage elasticity is always larger than the net-of-tax rate elasticity, and the difference—the wage-tax elasticity wedge—is statistically significant for all three types. This wedge is meaningfully large, with the Frisch net-of-tax rate elasticity being 30 to 50 percent smaller than the corresponding wage elasticity.

Intuitively, we should expect the wage-tax elasticity wedge to be larger if the respondent has a better opinion of the government and its programs. Table 6 provides some evidence that this is true. The table examines five categories of opinion questions that we included in our survey (two slightly different questions each, by design): satisfaction with the goverment and the way it spends its tax money, whether the government has an important role to play, trust in the government, whether the respondent thought she benefitted from government policies, and whether the existing redistribution system is fair. We standardize the questions to be on the same scale, take the average score of the questions in each category, and dichotomize this average. The choice variable in the logit specification is whether the respondent preferred the scenario over the other. All heterogeneity coefficients are positive as expected, although the difference is small and not statistically significant for the importance of the government, and only marginally significant at the 10 percent level for trust in government.[21]

4.2 Robustness

Table 5 shows robustness of our estimated Frisch wage-tax elasticity wedge to a variety of specification changes. Row 0 of the table repeats our main estimates in Panel C of Table 4, and all subsequent rows show a deviation in specification from row 0.

The first four rows examine robustness to varying the specification of the utility function in Equation (7). In row 1, we use a more flexible social preferences term $v(g)$ by increasing the polynomial order of g from three to five. In row 2, we allow tax-funded government

[20] In particular, we asked respondents to make their choices from a situation of nonemployment, which means that respondents should be abstracting away from constraints on hours (e.g. imposed by employers, or because rearranging daily schedules is costly). Since hours constraints attenuate intensive margin Frisch elasticities estimated based on real market situations (see e.g. Chetty et al. 2011), we should expect our estimates in column 3 of the table to be larger than corresponding estimates from the literature.

[21] We also examined heterogeneity by political affiliation, but are underpowered to pick up significant effects.

expenditure to be complements or substitutes to consumption and leisure by including the interactions of g with $\log(c)$ and $\log\left(\bar{L}-h\right)$. Another commonly-used parameterization of a discrete choice labor supply model is the quadratic utility function: $\tilde{u}(c,h) = \beta_c c + \beta_h\left(\bar{L}-h\right)+\beta_{cc}c^2+\beta_{hh}\left(\bar{L}-h\right)^2+\beta_{ch}c\left(\bar{L}-h\right)$, although Löffler et al. (2018) and Elder et al. (2023) report that their estimated elasticities are not sensitive to the choice of translog versus quadratic utility. Row 3 checks robustness of the wage-tax elasticity wedge by replacing the terms involving $\log(c)$ and $\log\left(\bar{L}-h\right)$ in Equation (7) with $\tilde{u}(c,h)$. Previous papers also frequently model parameter heterogeneity in demographics, and estimate elasticities by marital status and sex. In row 4, we allow β_c and β_h to vary with sex, marital status (married or not), the interaction of the above two, a quadratic in log age, a quadratic in the number of children in the family (top-coded at 9), and education (any college experience and whether the respondent has a four-year college degree or higher).[22] In general, we find that the estimated wage-tax elasticity wedge is robust to how we specify the utility function.

Our main specification balances survey burden and statistical power by collecting consumption information in only pairs 5 to 8, and then imputing it for all other pairs using the observed individual-level income-consumption relationship. Row 5 shows that our estimates are not sensitive to the imputation procedure by directly using the non-imputed consumption information (and restricting the estimation sample to pairs 5 to 8). Additionally, row 6 shows that our results are qualitatively similar if we use empirical Bayes prediction as an alternative imputation method. Here, we treat the individual-specific slope of log consumption on log take-home earnings (or the elasticity of consumption with respect to income) as random instead of fixed and shrink their estimates towards the average in the sample, using the mixed-model estimation approach described in Rabe-Hesketh and Skrondal (2012).[23]

Next, to account for the expected indirect utility term in Equation 1, our main specification controls for a quadratic in the log of the sum of the individual's assets, spouse's income, and the amount of income not consumed (or money borrowed, if applicable) of the scenario. Row 7 shows that the wage-tax elasticity wedge is even larger if we exclude this term.

By preference transitivity, a respondent who would work in one scenario of a pair but not the other should also prefer the former in the "choice between scenarios" question. When designing the survey, we worried that allowing the inconsistent choices might lead respondents

[22] In Appendix Table A2, we use the estimates from this model to simulate the elasticities for the four combinations of sex and marital status. Unlike papers that estimate labor supply elasticities using market data, we do not find larger labor supply elasticities for women than for men. This suggests that institutional setting or market friction drives the difference across gender.

[23] We also allow the slope to vary with demographic characteristics by including the interactions between take-home earnings and each of sex, marital status, the interaction of sex and marital status, education, a quadratic in log age, and a quadratic in the number of children in the family, in the fixed-effects portion of the model.

(who noticed) to think that the survey was badly designed and lose motivation. To avoid this, we included a non-intrusive prompt about the inconsistency whenever it was made. A natural worry is whether this leads to experimenter effects (e.g. the respondent may want to avoid seeing the text even though it does not interrupt the flow of his choices). To address this, we kept track of when the prompt was observed. In our final sample, 19% of respondents saw the prompt at least once, with the median first-observation of the prompt at pair 4, and 4% of respondents still made the intransitive choice despite the prompt. In row 8, we show that estimates are similar if we use only pairs before any observation of the prompt, suggesting that any possible experimenter effects or intransitivity does not matter for the wedge.[24]

Alternatively, making the intransitive choice might be a sign that attention was waning. Assuming that the intransitive choice was not made on purpose, the 19% of respondents who ever made the intransitive choice are less careful respondents whom we might want to exclude. In row 9, we show that estimates are robust to this exclusion.

We kept track of various measures of response quality throughout the survey. Row 10 reports that our results are robust to using a quality sample that excludes respondents who were inattentive, impatient, or responded too quickly. Specifically, we exclude respondents who were inattentive in either of our two attention check questions described in Section 3.2, clicked "next slide" more than twenty times in our instructions slides,[25] or answered any of the overall, vignette, and government expenditure liking questions in less than half the median time taken by all respondents.[26]

Our main estimates are based on the means of the simulated values. In Row 11, we check that the estimated wedge is not driven by outliers by reporting the median simulated wedge instead of the mean. (We boostrap the standard errors for inference.) The estimated median wedges are similar in magnitude to our main estimates.

[24]The logit functional form used in this paper imposes an independence of irrelevant alternatives (IIA) assumption, and papers that estimate discrete choice labor supply models frequently relax this by including random parameter slopes (Van Soest 1995; Euwals and Van Soest 1999; Löffler et al. 2018). Since IIA is related to transitivity through the weak axiom of revealed preference (see e.g. Peters and Protopapas 2021), the robustness of our result suggests that the IIA assumption is benign. This is consistent with Elder et al. (2023) who report that their results are similar whether random slopes were included or not, despite random slopes being much more computationally costly. Because of this, and because the primary goal of this paper is not in investigating heterogeneity in social preferences (e.g. as in Kosar et al. 2019; Burbano et al. forthcoming), we do not use a random slopes specification.

[25]Only ten clicks were needed. We also made this obvious by graying out the button to advance when instructions were still being displayed, and the superfluous clicks would not have produced a response from the system in any case.

[26]Reponses that are too quick are generally associated with poor response quality (Zhang and Conrad 2014; Greszki et al. 2015; Leiner 2019). While defining "too fast" is difficult, setting a threshold relative to the median is frequently done in practice (see e.g. Greszki et al. 2015).

As mentioned in Section 3.1, Table 1 shows minor imbalances along some demographic characteristics in our sample compared to the US population. To address this, we construct post-stratification weights using a raking procedure based on all the variables listed in the table (DeBell and Krosnick 2009). We apply this weight in the last row of Table 5. The estimated wedges remain similar with larger standard errors.

4.3 The importance of the program that taxes fund

Our main estimates in Section 4.1 show that there is a wedge between the responses of labor supply to wages and taxes. In this section, we provide evidence that social preferences towards tax-funded government expenditure plays a role in driving this wedge. Intuitively, if such social preferences matter, we should expect the wage-tax elasticity wedge to be larger if taxes fund specific programs that respondents like. (In a similar vein, previous papers have found that the programs that taxes fund matters for tax compliance [Alm et al. 1993; Hall and Preston 2000; Giaccobasso et al. 2022; Falsetta et al. 2023].) We reserved the last two pairs of scenarios in our survey to investigate this.[27]

These scenarios showed respondents randomly-drawn wages and work hours as before, but the randomly-drawn tax rates were now earmarked for a specific program, for example, funding education. (Appendix Figure A3 shows the instructions and layout.) Each pair of scenarios received the same program, and the programs differed across pairs 9 and 10 for greater within-person variation.[28] The program-specific tax rate was drawn from a more restricted grid compared to before (2, 4, 6, 8, or 10 percent), and to maintain continuity with previous pairs, respondents were shown a general tax rate set at their actual (simulated) tax rate rounded to the nearest even number, or 10 percent, whichever was larger. (This is the center of the tax rate grid that values were drawn from in previous scenarios.)

Near the end of the survey (six screens later), we asked respondents how much they liked or disliked their tax money being used to fund five different programs, including these

[27] We do not claim that tax-funded government expenditure is the only channel through which social preferences induce the wedge. While our vignette experiment shut down job- and environment-specific reasons like boss factors or promotion prospects (because we told respondents to assume that the jobs are identical to actual situations), social preferences in the workplace related to reciprocity to the employer's (un)generosity (Krueger and Mas 2004; Gneezy and List 2006; Mas 2006; Kube et al. 2012, 2013; DellaVigna et al. 2022) or pay inequality and worker's morale (Cohn et al. 2014; Bracha et al. 2015; Breza et al. 2017; Huet-Vaughn 2017; Dube et al. 2019; Cullen and Perez-Truglia 2022) might also play a role. Our experiment also shut down many market frictions—in particular, both wages and taxes were designed to be equally salient—which might also contribute to the wedge in practice. We leave the exploration of other mechanisms to future work.

[28] We elected not to show two programs on the same screen (i.e. in the same pair) to avoid experimenter effects. Assuming respondents focused on the decision between the two scenarios on the screen, different programs on different screens would avoid this.

two. The five programs were randomly drawn from a list of ten US federal budget functions—we did not elicit opinions on the full list to minimize respondent burden. Liking was measured on a 5-point Likert scale, with the two extremes corresponding to "dislike a lot" and "like a lot" (we randomized the scale direction between respondents). To ensure that respondents understood the programs, we included a short description in both the last two pairs of vignettes and the liking questions. Appendix Figure A4 plots the mean liking of the ten programs that we use, additionally broken down by political affiliation. On average, respondents liked their tax monies being used for retirement-related programs the most (social security and Medicare) and international affairs (which includes international aid) the least, with Democrats liking tax-funded expenditure more than Republicans on all programs except for national defense. This pattern is broadly consistent with what one might expect and suggests that respondents were careful in their responses.

Analysis of the last two pairs of our vignettes is now more complicated because changes in pre-tax earnings also induce variation in the amount of general taxes paid. Combined with our decision to draw the additional tax rate from a more limited range—a large program-specific tax rate would seem odd to the respondent—modeling the general and specific tax rate as one simple tax rate does not provide enough statistical power for analysis. Instead, we model them separately in the utility function:

$$V\left(a_i, w_i, \tau, \sigma, N_i\right) = \max_{c_i, a_i', h_i} U\left(c_i, h_i, g_i, s_i\right) + \beta E\left[V\left(a_i', w_i', \tau', \sigma', N_i'\right)\right] \tag{9}$$

$$\text{s.t. } c_i + \frac{1}{1+r} a_i' = a_i + \left(1 - \tau - \sigma\right) w_i h_i + N_i, \tag{10}$$

$$g_i = \tau w_i h_i, \tag{11}$$

$$s_i = \sigma w_i h_i, \tag{12}$$

where the new terms s_i and σ stand for the program-specific taxes paid and tax rate respectively. (Appendix B.3 shows the functional forms for the elasticities based on this model.) We parameterize utility with

$$\begin{aligned} U\left(c_{itj}, h_{itj}, g_{itj}, s_{itj}\right) = {} & \beta_c \log c_{itj} + \beta_h \log\left(\bar{L} - h_{itj}\right) + \beta_{cc} \left(\log c_{itj}\right)^2 \\ & + \beta_{hh} \left(\log\left(\bar{L} - h_{itj}\right)\right)^2 + \beta_{ch} \log c_{itj} \log\left(\bar{L} - h_{itj}\right) \\ & + \beta_g g_{itj} + \beta_s s_{itj} + \beta_{gs} g_{itj} s_{itj} + \xi_{itj}, \end{aligned} \tag{13}$$

$$\beta_l = \alpha_{0l} + \alpha_{1l} K_{it}, \quad l \in \{c, h, cc, hh, ch, g, s, gs\}, \tag{14}$$

where K_{it} is respondent i's liking score for the program shown in pair t, modeled as a

continuous variable.

Table 7 reports the relationship between the wage-tax elasticity wedge and how much the respondent likes the program that taxes fund. For better precision, we use the quality sample that excluded respondents who were inattentive, impatient, or responded too quickly from Section 4.2. As the table shows, the wage-tax elasticity wedge is larger if taxes go into funding programs that respondents like more. This is again consistent with the existence of government-related social preferences.

5 Government-related social preferences and the ETI

The vignette experiment allows us to experimentally manipulate wages and tax rates to obtain wage and net-of-tax rate elasticities for the same individuals. Its main drawback is that the result is based on stated preferences, and we have no way of verifying if the respondents would make the same choices in the real labor market.[29] In this section, we examine correlational evidence between proxies for government-related social preferences and estimates of the ETI with respect to the net-of-tax rate collected by Neisser (2021).

We select the ETI meta-analysis by Neisser (2021) for several reasons. First, as mentioned in the introduction, ETI estimates should also be affected by government-related social preferences since they are always estimated using tax variation. Second, this literature has a sizable body of estimates that are comparable to one another in terms of methodology. In contrast, the more mature labor supply elasticities literature has much greater variability in terms of methods and modeling decisions made, which makes comparability more difficult, and in some cases it is not clear if the variation comes from wage or tax changes.[30]

In total, Neisser (2021) examined 61 studies covering 17 countries. We drop Israel (one study) from the list because it participated in only one wave of the WVS, and its inclusion of WVS questions was only partial in that wave.[31] We examine seven government-related social preferences proxies: trust in the government, political parties, parliament, and civil service; whether income should be made more equal (to proxy for a desire for redistribution); whether the government should increase its ownership of businesses (more state ownership implies a more important government), and national pride (a general state-related social preference). Question wording of the proxies are similar in the WVS and EVS, and all proxies appear in

[29] That said, estimates based on stated preferences have been shown to correspond to actual market decisions in several other settings (Mas and Pallais 2017; Wiswall and Zafar 2018; Parker and Souleles 2019; Maestas et al. 2023).

[30] Meta-analyses of the conventional labor supply elasticity include Evers et al. (2008), Bargain and Peichl (2016), and Elminejada et al. (forthcoming).

[31] Results including Israel are qualitatively similar.

multiple waves. We collapse the proxies to the country-wave level, and match to the nearest WVS or EVS year based on the mean study year compiled in Neisser's replication kit.

Empirically, we follow Neisser's main specification in column 6 of her Tables 2 and 3 with three modifications. First, we include the proxy for government-related social preferences. Second, we exclude her country group variable in order to exploit all variation across countries. Third, we combine both before- and after-deductions elasticities for more statistical power.[32] Note that using Neisser's specification means that we already control for a large number of study-related covariates, like whether the study uses an instrumental variable technique and the decade that the tax policy was changed.

Table 8 Panel A shows results based on the above specification. In general, the proxies for government-related social preferences are negatively correlated with the ETI. Two proxies, the level of confidence in the civil service and the desire for equality in incomes, are not statistically significant, with the former significant at the 10% level. Significance level aside, the negative correlations suggest that when people have more positive opinions on the functions of the government, the response of taxable income to increases in the tax rate becomes smaller.

Panel B of the table shows robustness checks, with each row showing a change in specification compared to that in Panel A. In the first and second rows, we restrict the sample to before- and after-deductions elasticities respectively, similar to how Neisser (2021) presents her analysis. Estimates are qualitatively similar, with the correlations for before-deductions elasticities being smaller in magnitudes, consistent with after-deductions elasticities having more possible margins of responses (Slemrod and Kopczuk 2002; Kopczuk 2005; Neisser 2021). Next, Neisser investigated seven other correlates of the ETI that have been noted in the literature. In row 3, we control for her correlates except the income shares of the top 1 percent. (We find that the top 1 percent income share is correlated with the top 10 percent income share, so much so that standard errors increase significantly when both are included.) In row 4, we include the country-group fixed effects from Neisser's specification, and in row 5, we additionally include time fixed effects, where time corresponds to the waves of the WVS. The correlations between the ETI and the proxies are robust to the additional variables. Finally, in row 6, we control for the share of public spending on the provision of child care, preschool, and elderly care. This variable is from Kleven (2014), who notes that these public goods are complementary to labor supply, and hence the mechanism could be

[32] A major point of Neisser (2021) is that the after-deductions ETI is larger and more sensitive with respect to the estimation method than the before-deductions ETI. The difference between the two, however, is not relevant for our purposes. Hence, we analyze both ETI types as one sample, include a dummy for whether the ETI is a before- or after-deductions elasticity, and interact this dummy with all variables except the social preferences proxy.

due to the way that taxes are used rather than government-related social preferences. The correlations are again robust to the inclusion of this variable.

6 Conclusion

This paper studies the implications of social preferences on the response of labor supply to wages and taxes. Using a vignette experiment, we show that when social preferences are incorporated in the canonical labor supply model, estimates of the labor supply elasticity that exploit exogenous variation in wages are larger than those that exploit variation in tax rates. This wage-tax elasticity wedge is meaningfully large—in our main estimate, the frictionless Frisch wage elasticity is 1.5 times as large as the corresponding net-of-tax rate elasticity. Furthermore, the wedge is larger if the respondent has a better opinion of the government, or if the taxes are earmarked for funding a program that the respondent likes better. In a complementary analysis, we find that estimates of the ETI with respect to the net-of-tax rates are correlated with proxies for government-related social preferences, consistent with people exhibiting social preferences when faced with real life labor market decisions.

Our results have implications for how we use estimates of the labor supply elasticity from the literature in model calibration. In particular, macroeconomic models frequently calibrate a labor supply elasticity parameter to model responses to wage changes over the real business cycle. If a net-of-tax rate elasticity is used (e.g. by using estimates based on tax holiday natural experiments), the calibrated parameter would be too small, with implications for the possible recommendations made by such models.

On the other hand, the profession is likely already using the correct elasticity when assessing tax changes—most quasi-experimental estimates tend to be net-of-tax rate elasticities, and since Feldstein (1995), they tend to be (rightly) ETI estimates. That said, results of models evaluating tax changes can still change subtly once we allow people to have government-related social preferences. In Appendix B.4, we replicate the main arguments in Piketty and Saez (2013) to draw two conclusions about optimal taxation. First, admitting government-related social preferences into individuals' utility functions naturally changes the social welfare function, and if social preferences are positive as our results suggest, optimal tax rates would be higher relative to a world without government-related social preferences. Intuitively, knowing that people have preferences on taxation beyond its immediate impact on disposable income means that the government can (and should) raise taxes. Second, heterogeneity in social preferences would naturally affect the optimal nonlinear tax rate. We leave further detailed explorations along this line for future work.

Lastly, our findings that the wedge (and the ETI) vary by opinion of the government and programs that taxes fund has relevance for tax policy. The tax morale literature has emphasized that attitudes towards government or programs are important for tax compliance (Cowell and Gordon 1988; Alm et al. 1993; Hall and Preston 2000; Cullen et al. 2021; Giaccobasso et al. 2022; Falsetta et al. 2023). Our results show that such buy-in towards how the government spends its money is important for reducing the distortionary effect of taxes on labor supply as well. This provides additional justification for policies on government transparency and accountability—for example, communications about how tax monies are being used, or reminders that tax monies fund services. To the extent that people have more positive attitudes towards local rather than geographically larger communities, our results also lend some support for fiscal devolution by pointing out a benefit often missed in the discussion (see e.g. Martinez-Vazquez et al. 2017). More speculatively, our findings also suggests that tax collection practices that engender negative attitudes (e.g. differential opportunities to reduce or avoid tax liabilities, or distrusting tax systems described in Frey [1997]) are particularly undesirable, because of the additional consequences for labor supply.

References

Alesina, Alberto and George-Marios Angeletos, "Fairness and redistribution," *American Economic Review*, 2005, *95* (4), 960–980.

_ **and Paola Giuliano**, "Preferences for redistribution," in "Handbook of Social Economics," Vol. 1, Elsevier, 2011, pp. 93–131.

_ **, Matteo F Ferroni, and Stefanie Stantcheva**, "Perceptions of racial gaps, their causes, and ways to reduce them," Working Paper 29245, National Bureau of Economic Research 2021.

Alm, James, "What motivates tax compliance?," *Journal of Economic Surveys*, 2019, *33* (2), 353–388.

_ **, Betty R Jackson, and Michael McKee**, "Fiscal exchange, collective decision institutions, and tax compliance," *Journal of Economic Behavior and Organization*, 1993, *22* (3), 285–303.

Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden, "Global evidence on the selfish rich inequality hypothesis," *Proceedings of the National Academy of Sciences*, 2022, *119* (3), e2109690119.

Altonji, Joseph G, "Intertemporal substitution in labor supply: Evidence from micro data," *Journal of Political Economy*, 1986, *94* (3, Part 2), S176–S215.

Andreoni, James, "Giving with impure altruism: Applications to charity and Ricardian equivalence," *Journal of Political Economy*, 1989, *97* (6), 1447–1458.

_ , "Impure altruism and donations to public goods: A theory of warm-glow giving," *The Economic Journal*, 1990, *100* (401), 464–477.

_ **and A Abigail Payne**, "Charitable giving," in "Handbook of Public Economics," Vol. 5, Elsevier, 2013, pp. 1–50.

_ **, Brian Erard, and Jonathan Feinstein**, "Tax compliance," *Journal of Economic Literature*, 1998, *36* (2), 818–860.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin, "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 1996, *91* (434), 444–455.

Baker, Scott R., "Debt and the response to household income shocks: Validation and application of linked financial account data," *Journal of Political Economy*, 2018, *126* (4), 1504–1557.

Bargain, Olivier and Andreas Peichl, "Own-wage labor supply elasticities: variation across time and estimation methods," *IZA Journal of Labor Economics*, 2016, *5* (1), 1–31.

Bauer, Thomas K, Julia Bredtmann, and Christoph M Schmidt, "Time vs. money—The supply of voluntary labor and charitable donations across Europe," *European Journal of Political Economy*, 2013, *32*, 80–94.

Besley, Timothy, Anders Jensen, and Torsten Persson, "Norms, enforcement, and tax evasion," *The Review of Economics and Statistics*, 2023, *105* (4), 998–1007.

Bianchi, Marco, Bjorn R Gudmundsson, and Gylfi Zoega, "Iceland's natural experiment in supply-side economics," *American Economic Review*, 2001, *91* (5), 1564–1579.

Blau, Francine D and Lawrence M Kahn, "Changes in the labor supply behavior of married women: 1980–2000," *Journal of Labor economics*, 2007, *25* (3), 393–438.

Blumkin, Tomer, Bradley J Ruffle, and Yosef Ganun, "Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods," *European Economic Review*, 2012, *56* (6), 1200–1219.

Blundell, Richard, Alan Duncan, and Costas Meghir, "Estimating labor supply responses using tax reforms," *Econometrica*, 1998, pp. 827–861.

_ **and Thomas MaCurdy**, "Labor supply: A review of alternative approaches," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3, Amsterdam, The Netherlands: Elsevier, 1999, chapter 27, pp. 1559–1695.

_ **, Monica Costa Dias, Costas Meghir, and Jonathan Shaw**, "Female labor supply, human capital, and welfare reform," *Econometrica*, 2016, *84* (5), 1705–1753.

Borjas, George J, "The relationship between wages and weekly hours of work: The role of division bias," *Journal of Human Resources*, 1980, *15* (3), 409–423.

Bozio, Antoine, Thomas Breda, and Julien Grenet, "Does Tax-Benefit Linkage Matter for the Incidence of Social Security Contributions?," Discussion Paper 12502, IZA 2019. Retrieved May 12, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3435377.

Bracha, Anat, Uri Gneezy, and George Loewenstein, "Relative Pay and Labor Supply," *Journal of Labor Economics*, 2015, *33* (2), 297–315.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 10 2017, *133* (2), 611–663.

Burbano, Vanessa, Nicolas Padilla, and Stephan Meier, "Gender differences in preferences for meaning at work," *American Economic Journal: Economic Policy*, forthcoming.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor supply of New York City cabdrivers: One day at a time," *The Quarterly Journal of Economics*, 1997, *112* (2), 407–441.

Carpenter, Jeffrey, "The shape of warm glow: Field experimental evidence from a fundraiser," *Journal of Economic Behavior and Organization*, 2021, *191*, 555–574.

Chang, Yongsung and Sun-Bin Kim, "From individual to aggregate labor supply: A quantitative analysis based on a heterogeneous agent macroeconomy," *International Economic Review*, 2006, *47* (1), 1–27.

Charness, Gary and Matthew Rabin, "Understanding social preferences with simple tests," *The quarterly journal of economics*, 2002, *117* (3), 817–869.

Chen, M Keith, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen, "The value of flexible work: Evidence from Uber drivers," *Journal of political economy*, 2019, *127* (6), 2735–2794.

Chetty, Raj, "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 2012, *80* (3), 969–1018.

_ , Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

_ , John N Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records," *The Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

Cingl, Lubomír, Tomáš Lichard, and Tomáš Miklánek, "Tax designation effects on compliance: An online experiment with taxpayers," *Journal of Economic Behavior and Organization*, 2023, *214*, 615–633.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider, "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association*, 08 2014, *12* (4), 877–898.

Cowell, Frank A and James P F Gordon, "Unwillingness to pay: Tax evasion and public good provision," *Journal of Public Economics*, 1988, *36* (3), 305–321.

Cullen, Julie Berry, Nicholas Turner, and Ebonya Washington, "Political alignment, attitudes toward government, and tax evasion," *American Economic Journal: Economic Policy*, 2021, *13* (3), 135–166.

Cullen, Zoë and Ricardo Perez-Truglia, "How much does your boss make? The effects of salary comparisons," *Journal of Political Economy*, 2022, *130* (3), 766–822.

DeBell, Matthew and Jon A Krosnick, "Computing weights for American National Election Study Survey data," ANES Technical Report series, no. nes012427, American National Election Studies, Ann Arbor, MI, and Palo Alto, CA 2009.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, "Testing for altruism and social pressure in charitable giving," *The Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

_ , John A. List, Ulrike Malmendier, and Gautam Rao, "Estimating Social Preferences and Gift Exchange at Work," *American Economic Review*, March 2022, *112* (3), 1038–74.

Diamond, Peter, "Optimal tax treatment of private contributions for public goods with and without warm glow preferences," *Journal of Public Economics*, 2006, *90* (4-5), 897–919.

Djanali, Iwan and Damien Sheehan-Connor, "Tax affinity hypothesis: Do we really hate paying taxes?," *Journal of Economic Psychology*, 2012, *33* (4), 758–775.

Doerrenberg, Philipp and Andreas Peichl, "Tax Morale and the Role of Social Norms and Reciprocity—Evidence from a Randomized Survey Experiment," *FinanzArchiv*, 2022, *78* (1-2), 44–86. Articles.

Doherty, Carroll, Jocelyn Kiley, Alec Tyson, and Bridget Jameson, "Beyond distrust: How Americans view their government," Technical Report, Pew Research Center 2015.

Domeij, David and Martin Floden, "The labor-supply elasticity and borrowing constraints: Why estimates are biased," *Review of Economic Dynamics*, 2006, *9* (2), 242–262.

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard, "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior," *American Economic Review*, February 2019, *109* (2), 620–63.

Durante, Ruben, Louis Putterman, and Joël Van der Weele, "Preferences for redistribution and perception of fairness: An experimental study," *Journal of the European Economic Association*, 2014, *12* (4), 1059–1086.

Eissa, Nada, "Taxation and labor supply of married women: The Tax Reform Act of 1986 as a natural experiment," Working Paper No. 5023, National Bureau of Economic Research 1995.

_ **and Hilary Williamson Hoynes**, "Taxes and the labor market participation of married couples: The earned income tax credit," *Journal of Public Economics*, 2004, *88* (9-10), 1931–1958.

Elder, Todd E, Steven Haider, and Cody Orr, "The evolution of the wage elasticity of labor supply over time," Discussion Paper 16393, IZA 2023. Retrieved Nov 4, 2023, from https://ssrn.com/abstract=4547697.

Elminejada, Ali, Tomas Havraneka, Roman Horvatha, and Zuzana Irsovaa, "Intertemporal substitution in labor supply: A meta-analysis," *Review of Economic Dynamics*, forthcoming.

Eriksson, Tor and Nicolai Kristensen, "Wages or fringes? Some evidence on trade-offs and sorting," *Journal of Labor Economics*, 2014, *32* (4), 899–928.

Esslinger, Anna, Katharina Pfeil, and Lars P Feld, "Asymmetric Labor Supply Responses to Taxation," Working Paper No. 11317, CESifo 2024.

Euwals, Rob and Arthur Van Soest, "Desired and actual labour supply of unmarried men and women in the Netherlands," *Labour Economics*, 1999, *6* (1), 95–118.

Evers, Michiel, Ruud De Mooij, and Daniel Van Vuuren, "The wage elasticity of labour supply: A synthesis of empirical estimates," *De Economist*, 2008, *156*, 25–43.

EVS, "EVS Trend File 1981-2017. ZA7503 Data file Version 3.0.0," 2022. GESIS Data Archive, Cologne. https://doi.org/10.4232/1.14021.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *The Review of Economic Studies*, 2016, *83* (2), 587–628.

Falsetta, Diana, Jennifer K Schafer, and George T Tsakumis, "How government spending impacts tax compliance," *Journal of Business Ethics*, 2023, pp. 1–18.

Farber, Henry S, "Is tomorrow another day? The labor supply of New York City cabdrivers," *Journal of political Economy*, 2005, *113* (1), 46–82.

Fehr, Ernst and Klaus M Schmidt, "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

_ **and Lorenz Goette**, "Do workers work more if wages are high? Evidence from a randomized field experiment," *American Economic Review*, 2007, *97* (1), 298–317.

Feldstein, Martin, "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act," *Journal of Political Economy*, 1995, *103* (3), 551–572.

Finkelstein, Amy, "E-ztax: Tax Salience and Tax Rates," *The Quarterly Journal of Economics*, 08 2009, *124* (3), 969–1010.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry, "Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]," 2022. Minneapolis, MN: IPUMS, Retrieved November 7, 2022, from https://doi.org/10.18128/D030.V10.0.

Folke, Olle and Johanna Rickne, "Sexual harassment and gender inequality in the labor market," *The Quarterly Journal of Economics*, 2022, *137* (4), 2163–2212.

Freeman, Richard B, "Working for nothing: The supply of volunteer labor," *Journal of Labor Economics*, 1997, *15* (1, Part 2), S140–S166.

Frey, Bruno S, "Tertium datur: Pricing, regulating and intrinsic motivation," *Kyklos*, 1992, *45* (2), 161–184.

_ , "A constitution for knaves crowds out civic virtues," *The Economic Journal*, 1997, *107* (443), 1043–1053.

Gallup, "Party Affiliation," 2023. Retrieved October 29, 2023, from https://news.gallup.com/poll/15370/party-affiliation.aspx.

Gelber, Alexander M, "Taxation and the earnings of husbands and wives: evidence from Sweden," *Review of Economics and Statistics*, 2014, *96* (2), 287–305.

Giaccobasso, Matias, Brad C Nathan, Ricardo Perez-Truglia, and Alejandro Zentner, "Where do my tax dollars go? Tax morale effects of perceived government spending," Working Paper 29789, National Bureau of Economic Research 2022.

Giné, Xavier, Monica Martinez-Bravo, and Marian Vidal-Fernández, "Are labor supply decisions consistent with neoclassical preferences? Evidence from Indian boat owners," *Journal of Economic Behavior & Organization*, 2017, *142*, 331–347.

Gneezy, Uri and John A List, "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

Goodrich, Brittney, Marieke Fenton, Jerrod Penn, John Bovay, and Travis Mountain, "Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys," *Applied Economic Perspectives and Policy*, 2023, *45* (2), 762–784.

Greszki, Robert, Marco Meyer, and Harald Schoen, "Exploring the effects of removing "too fast" responses and respondents from web surveys," *Public Opinion Quarterly*, 2015, *79* (2), 471–503.

Gruber, Jonathan, "The incidence of payroll taxation: Evidence from Chile," *Journal of Labor Economics*, 1997, *15* (S3), S72–S101.

Hall, John and Ian Preston, "Tax price effects on attitudes to hypothecated tax increases," *Journal of Public Economics*, 2000, *75* (3), 417–438.

Hotz, V Joseph, Finn E Kydland, and Guilherme L Sedlacek, "Intertemporal preferences and labor supply," *Econometrica: Journal of the Econometric Society*, 1988, pp. 335–360.

Huber, John D. and Piero Stanig, "Church-state separation and redistribution," *Journal of Public Economics*, 2011, *95* (7), 828–836.

Huet-Vaughn, Emiliano, "Do Social Comparisons Motivate Workers? A Field Experiment on Relative Earnings, Labor Supply and the Inhibitory Effect of Pay Inequality," Working Paper, Middlebury College 2017.

Hvidberg, Kristoffer B, Claus T Kreiner, and Stefanie Stantcheva, "Social positions and fairness views on inequality," *The Review of Economic Studies*, 03 2023, p. rdad019.

Imai, Susumu and Michael P Keane, "Intertemporal labor supply and human capital accumulation," *International Economic Review*, 2004, *45* (2), 601–641.

Imbens, Guido W. and Joshua D. Angrist, "Identification and estimation of local average treatment effects," *Econometrica*, 1994, *62* (2), 467–475.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen, "World Values Survey Trend File (1981-2022) Cross-National Data-Set. Data File Version 3.0.0.," 2022. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. https://doi.org/10.14281/18241.23.

Internal Revenue Service, "Statistics of income—2019 individual income tax returns," Publication 1304 (Rev. 12-2021), Department of the Treasury Internal Revenue Service, Washington, D.C. 2021. Retrieved September 27, 2022, from https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-returns-complete-report-publication-1304.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer, "Worker beliefs about outside options," Working Paper 29623, National Bureau of Economic Research 2023.

Karadja, Mounir, Johanna Mollerstrom, and David Seim, "Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution," *Review of Economics and Statistics*, 2017, *99* (2), 201–212.

Keane, Michael P, "Labor supply and taxes: A survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

_ , "Recent research on labor supply: Implications for tax and transfer policy," *Labour Economics*, 2022, *77*, 102026.

_ **and Nada Wasi**, "Labour supply: The roles of human capital and the extensive margin," *The Economic Journal*, 2016, *126* (592), 578–617.

Kessler, Judd B. and Michael I. Norton, "Tax aversion in labor supply," *Journal of Economic Behavior & Organization*, 2016, *124*, 15–28.

Kleven, Henrik Jacobsen, "How can Scandinavians tax so much?," *Journal of Economic Perspectives*, 2014, *28* (4), 77–98.

Konrad, Kai A and Salmai Qari, "The last refuge of a scoundrel? Patriotism and tax compliance," *Economica*, 2012, *79* (315), 516–533.

Kopczuk, Wojciech, "Tax bases, tax rates and the elasticity of reported income," *Journal of Public Economics*, 2005, *89* (11-12), 2093–2119.

Kosar, Gizem, Aysegul Sahin, Basit Zafar et al., "The work-leisure tradeoff: Identifying the heterogeneity," in "2019 Meeting Papers" number 410 Society for Economic Dynamics 2019.

Kroft, Kory, Jean-William Laliberté, René Leal-Vizcaíno, and Matthew J Notowidigdo, "Salience and taxation with imperfect competition," *Review of Economic Studies*, 2024, *91* (1), 403–437.

Krueger, Alan B. and Alexandre Mas, "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," *Journal of Political Economy*, 2004, *112* (2), 253–289.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe, "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, June 2012, *102* (4), 1644–62.

_ **,** _ **, and** _ , "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment," *Journal of the European Economic Association*, 08 2013, *11* (4), 853–870.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva, "How elastic are preferences for redistribution? Evidence from randomized survey experiments," *American Economic Review*, 2015, *105* (4), 1478–1508.

Lamberton, Cait, Jan-Emmanuel De Neve, and Michael I Norton, "The power of voice in stimulating morality: Eliciting taxpayer preferences increases tax compliance," *Journal of Consumer Psychology*, 2018, *28* (2), 310–328.

Leiner, Dominik Johannes, "Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys," *Survey Research Methods*, 2019, *13* (3), 229–248.

Leip, David, "Atlas of U.S. Presidential Elections," 2023. Retrieved October 29, 2023, from https://uselectionatlas.org/.

Levine, David K, "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, 1998, *1* (3), 593–622.

Lilley, Andrew and Robert Slonim, "The price of warm glow," *Journal of Public Economics*, 2014, *114*, 58–74.

Löffler, Max, Andreas Peichl, and Sebastian Siegloch, "The sensitivity of structural labor supply estimations to modeling assumptions," Discussion Paper 11425, IZA 2018. Retrieved Nov 18, 2023, from https://www.iza.org/publications/dp/11425/the-sensitivity-of-structural-labor-supply-estimations-to-modeling-assumptions.

Low, Hamish W, "Self-insurance in a life-cycle model of labour supply and savings," *Review of Economic Dynamics*, 2005, *8* (4), 945–975.

Luttmer, Erzo F P and Monica Singhal, "Culture, context, and the taste for redistribution," *American Economic Journal: Economic Policy*, 2011, *3* (1), 157–179.

Luttmer, Erzo FP and Monica Singhal, "Tax morale," *Journal of Economic Perspectives*, 2014, *28* (4), 149–168.

MaCurdy, Thomas E, "An empirical model of labor supply in a life-cycle setting," *Journal of Political Economy*, 1981, *89* (6), 1059–1085.

Maestas, Nicole, Kathleen J Mullen, David Powell, Till Von Wachter, and Jeffrey B Wenger, "The value of working conditions in the United States and implications for the structure of wages," *American Economic Review*, 2023, *113* (7), 2007–2047.

Martinez, Isabel Z, Emmanuel Saez, and Michael Siegenthaler, "Intertemporal labor supply substitution? Evidence from the Swiss income tax holidays," *American Economic Review*, 2021, *111* (2), 506–546.

Martinez-Vazquez, Jorge, Santiago Lago-Peñas, and Agnese Sacchi, "The impact of fiscal decentralization: A survey," *Journal of Economic Surveys*, 2017, *31* (4), 1095–1129.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *The Quarterly Journal of Economics*, 08 2006, *121* (3), 783–821.

_ **and Amanda Pallais**, "Valuing alternative work arrangements," *American Economic Review*, December 2017, *107* (12), 3722–3759.

_ **and** _ , "Labor supply and the value of non-work time: Experimental estimates from the field," *American Economic Review: Insights*, 2019, *1* (1), 111–126.

McFadden, Daniel, "Conditional logit analysis of qualitative choice behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*, New York, NY: Academic Press, 1974, chapter 4.

Mori, Tomoharu, Hirofumi Kurokawa, and Fumio Ohtake, "Labor supply reaction to wage cuts and tax increases: A real-effort experiment," *FinanzArchiv*, 2022, *78* (3), 1–16.

Mui, Preston and Benjamin Schoefer, "Reservation raises: The aggregate labour supply curve at the extensive margin," *Review of Economic Studies*, forthcoming.

Nathan, Brad C, Ricardo Perez-Truglia, and Alejandro Zentner, "My taxes are too darn high: Why do households protest their taxes?," Working Paper 27816, National Bureau of Economic Research 2021.

Neisser, Carina, "The elasticity of taxable income: A meta-regression analysis," *The Economic Journal*, 2021, *131* (640), 3365–3391.

Oettinger, Gerald S, "An empirical analysis of the daily labor supply of stadium vendors," *Journal of Political Economy*, 1999, *107* (2), 360–392.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie, "Why do people give? Testing pure and impure altruism," *American Economic Review*, 2017, *107* (11), 3617–3633.

Parker, Jonathan A and Nicholas S Souleles, "Reported effects versus revealed-preference estimates: Evidence from the propensity to spend tax rebates," *American Economic Review: Insights*, 2019, *1* (3), 273–290.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2022, p. 1.

Peters, Hans and Panos Protopapas, "Set and revealed preference axioms for multi-valued choice," *Theory and Decision*, 2021, *90*, 11–29.

Piketty, Thomas and Emmanuel Saez, "Optimal labor income taxation," in "Handbook of Public Economics," Vol. 5, Elsevier, 2013, pp. 391–474.

Pistaferri, Luigi, "Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply," *Journal of Labor Economics*, 2003, *21* (3), 729–754.

Rabe-Hesketh, Sophia and Anders Skrondal, *Multilevel and Longitudinal Modeling Using Stata, Volumes I and II*, College Station, TX: Stata Press, 2012.

Rick, Scott, Gabriele Paolacci, and Katherine Burson, "Income tax and the motivation to work," *Journal of Behavioral Decision Making*, 2018, *31* (5), 619–631.

Rogerson, Richard and Johanna Wallenius, "Micro and macro elasticities in a life cycle model with taxes," *Journal of Economic Theory*, 2009, *144* (6), 2277–2292.

Saez, Emmanuel, "The optimal treatment of tax expenditures," *Journal of Public Economics*, 2004, *88* (12), 2657–2684.

_ , Joel Slemrod, and Seth H Giertz, "The elasticity of taxable income with respect to marginal tax rates: A critical review," *Journal of Economic Literature*, 2012, *50* (1), 3–50.

Shaw, Kathryn L, "Life-cycle labor supply with human capital accumulation," *International Economic Review*, 1989, pp. 431–456.

Sigaard, Hans Schytte, "Estimating labor supply responses to Danish tax reforms," *Journal of Public Economics*, 2023, *224*, 104943.

Sigurdsson, Jósef, "Labor Supply Responses and Adjustment Frictions: A Tax-Free Year in Iceland," Working Paper 2019. Available at SSRN: https://ssrn.com/abstract=3278308.

Slemrod, Joel and Wojciech Kopczuk, "The optimal elasticity of taxable income," *Journal of Public Economics*, 2002, *84* (1), 91–112.

Soest, Arthur Van, "Structural models of family labor supply: A discrete choice approach," *Journal of Human Resources*, 1995, pp. 63–88.

Stafford, Tess M, "What do fishermen tell us that taxi drivers do not? An empirical investigation of labor supply," *Journal of Labor Economics*, 2015, *33* (3), 683–710.

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

_ , "Understanding of trade," Working Paper 30040, National Bureau of Economic Research 2022.

_ , "How to run surveys: A guide to creating your own identifying variation and revealing the invisible," *Annual Review of Economics*, 2023, *15*, 205–234.

Stefánsson, Arnaldur, "Labor supply response to a tax holiday: The take-home from a large and salient shock," Working Paper, Uppsala University 2019.

Summers, Lawrence H, "Some simple economics of mandated benefits," *The American Economic Review*, 1989, *79* (2), 177–183.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention variation and welfare: Theory and evidence from a tax salience experiment," *The Review of Economic Studies*, 2018, *85* (4), 2462–2496.

Unrath, Matthew, "Beating the clock: Using year-end changes to identify intensive margin labor supply responses to taxation," Working Paper, UC Berkeley 2020. Retrieved Nov 4, 2023, from https://www.mattunrath.com/research.

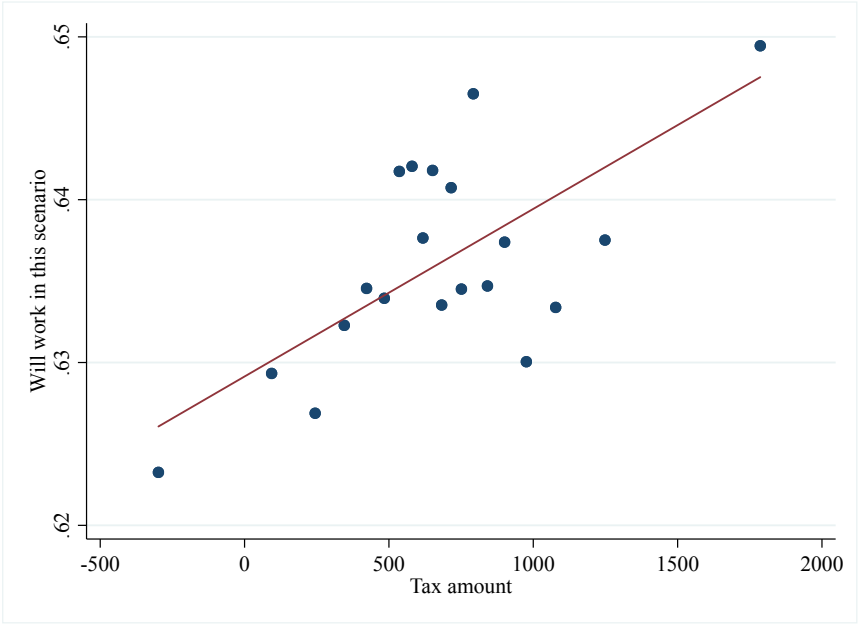
Wiswall, Matthew and Basit Zafar, "Preference for the workplace, investment in human capital, and gender," *The Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

Zhang, Chan and Frederick Conrad, "Speeding in web surveys: The tendency to answer very fast and its association with straightlining," *Survey Research Methods*, 2014, *8* (2), 127–135.

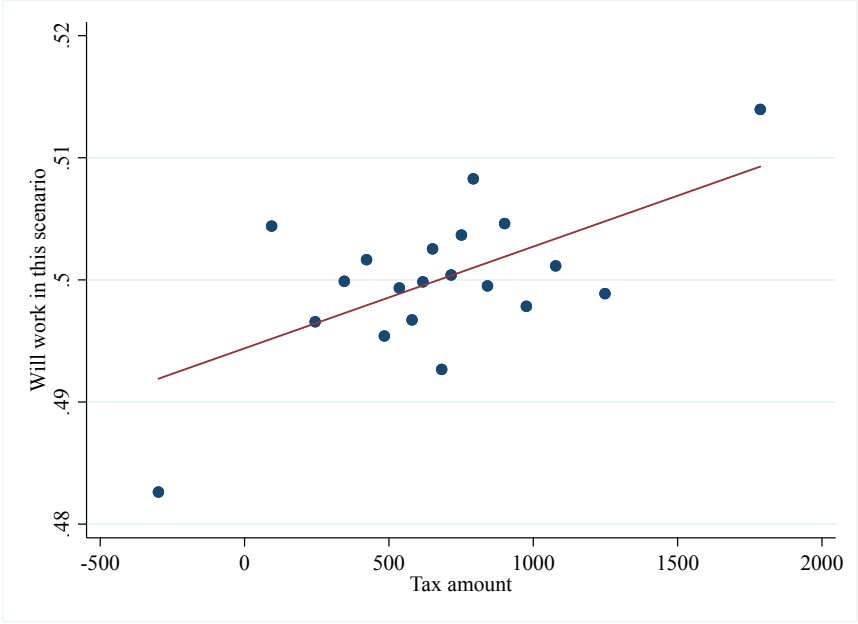
Ziliak, James P and Thomas J Kniesner, "The effect of income taxation on consumption and labor supply," *Journal of Labor Economics*, 2005, *23* (4), 769–796.

Figure 1: Binscatter plots of choices made against tax amount

(a) Choice of work or not



(b) Choice of whether to pick the scenario over the other



Notes: The sample comprises respondent-pair-scenarios with valid in-vignette consumption data, excluding respondents in the top and bottom 5 percent of actual work hours, consumption, and simulated tax rate. The two panels show binscatter plots of the choices made by respondents against the amount of taxes paid in the scenario. In Panel A, the choice made is whether to work or not in the scenario. In Panel B, the choice made is whether the respondent preferred the scenario over the other in the pair. Both binscatter plots partial out individual, pair, and option fixed effects, as well as centile of disposable income by centile of work hours fixed effects.

Table 1: Summary Statistics

	U.S. Population	Survey Sample
Male	0.52	0.51
25-34 years old	0.29	0.28
35-44 years old	0.27	0.26
45-54 years old	0.24	0.24
55+ years old	0.21	0.22
0-19,999 dollars	0.12	0.08
20,000-39,999 dollars	0.23	0.24
40,000-59,999 dollars	0.23	0.27
60,000-99,999 dollars	0.24	0.28
100,000+ dollars	0.18	0.13
Four-year college degree or more	0.43	0.60
High-school graduate or less	0.31	0.12
Married	0.61	0.52
White	0.76	0.82
Black	0.12	0.08
Asian	0.08	0.07
Others	0.04	0.04
Hispanic	0.20	0.09
Republican	0.30	0.27
Democrat	0.29	0.40
Independent	0.38	0.26
Voted for Biden in the 2020 presidential election	0.51	0.51
Voted for Trump in the 2020 presidential election	0.47	0.32
Sample Size		5,440

Notes: This table presents statistics for the overall U.S. population aged 25-64 who are employed and compares them to the characteristics of our survey respondents. National statistics on gender, age, income brackets, education, marital status, and race are from the IPUMS Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) data set for 2022 (Flood et al. 2022). National statistics on party affiliation for May 2023 are from Gallup (2023). Note that these statistics are based on surveys of adults 18 years and older and are not restricted to employed individuals. Presidential election results from 2020 are from Leip (2023).

Table 2: Reduced form evidence for social preferences

	Dependent variable: Will work in this scenario.			Dependent variable: Prefers this scenario over other.		
	Fifth-order polynomial in controls	FEs for interacted centiles of controls	FEs from (2), further allowing heterogeneous slopes within cells	Fifth-order polynomial in controls	FEs for interacted centiles of controls	FEs from (5), further allowing heterogeneous slopes within cells
	(1)	(2)	(3)	(4)	(5)	(6)
Tax paid ($1000)	0.021***	0.010***	0.014***	0.014***	0.0083**	0.014***
	(0.0045)	(0.0031)	(0.0034)	(0.0041)	(0.0036)	(0.0042)
Observations	87,037	86,978	86,978	87,040	86,981	86,981
Respondents	5,440	5,440	5,440	5,440	5,440	5,440

Notes: Standard errors clustered by respondents in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises respondent-pair-scenarios with valid in-vignette consumption data, excluding respondents in the top and bottom 5 percent of actual work hours, consumption, and simulated tax rate. All columns control for individual, pair, and scenario order fixed effects. Columns 1 and 4 control for all first- to fifth-order interaction terms involving disposable income and work hours. Columns 2 and 5 control non-parametrically for disposable income and work hours by including centile of disposable income by centile of work hours fixed effects. Columns 3 and 6 include the same fixed effects, and further control for the fixed effects interacted (separately) with disposable income and work hours.

Table 3: Estimates of utility parameters

	Dependent variable:			
	Will work in this scenario		Prefers this scenario over other	
	(1)	(2)	(3)	(4)
Log(Consumption)	2.26	0.47	1.14	-0.37
	(1.63)	(1.63)	(1.37)	(1.39)
Log(672 - Work hours)	219.3***	204.2***	143.5***	134.4***
	(20.7)	(20.8)	(13.5)	(13.6)
Log(Consumption) × Log(Consumption)	0.18***	0.17***	0.19***	0.18***
	(0.018)	(0.018)	(0.032)	(0.032)
Log(672 - Work hours) × Log(672 - Work hours)	-16.9***	-15.7***	-11.0***	-10.3***
	(1.63)	(1.64)	(1.07)	(1.07)
Log(Consumption) × Log(672 - Work hours)	-0.44*	-0.15	-0.29*	-0.055
	(0.25)	(0.25)	(0.18)	(0.18)
Part-time	-0.23***	-0.23***	-0.17***	-0.17***
	(0.056)	(0.057)	(0.036)	(0.036)
Tax paid ($1000)		0.89***		0.48***
		(0.089)		(0.070)
Tax paid ($1000) ^ 2		-0.17***		-0.064***
		(0.038)		(0.024)
Tax paid ($1000) ^ 3		0.0068**		0.0016**
		(0.0032)		(0.00078)
Observations	36,038	36,038	80,434	80,434
Respondents	4,671	4,671	5,381	5,381
$E(U_c) \times 1,000$	1.1***	1.0***	1.0***	0.99***
$E(U_G) \times 1,000$		0.70***		0.41***
$E(U_G)/E(U_c)$		0.69		0.41
$E(U_{GG}G) \times 1,000$		-0.17***		-0.069***
Log pseudo-likelihood	-8612.4	-8442.8	-21773.8	-21596.4

Notes: Standard errors clustered by respondents in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises respondent-pair-scenarios with valid in-vignette consumption data, excluding respondents in the top and bottom 5 percent of actual work hours, consumption, and simulated tax rate. All variables are on a per-month basis. Estimates are based on a conditional (on respondent-pair) logit model, controlling for option fixed effects, an indicator for whether the scenario's weekly work hours is below 35, and a squared polynomial term in $\log(assets_i + spouseincome_i + savings_{itj})$, where $assets_i$ is the wealth in dollars of the individual, $spouseincome_i$ is the individual's spouse's income (or zero if not applicable), and $savings_{itj}$ is the after-tax disposable income less consumption for the scenario. Consumption is the imputed consumption for the scenario, see Section 3.3 for details. Work hours are the work hours for the scenario, so that 672 (the number of hours in 4 weeks) is the amount of leisure time for the scenario. Tax paid is amount of taxes to be paid for the scenario, in thousand dollars. $E(U_c)$ and $E(U_g)$ are the derivatives of the utility function in Equation (7), computed using the estimated parameters in the table and the actual values reported (for non-tax variables) or simulated (for taxes paid and tax rates) by the respondent, averaged over respondents in the sample (one per respondent). Standard errors are computed using the delta method. $E(U_{gg}g)$ is the similarly-computed second-order derivative multiplied by the product of the simulated tax rate, reported hourly wage, and reported work hours of the respondent. $E(U_g)/E(U_c)$ is the ratio of the averaged derivatives in rows above. Where relevant, we multiply utility derivatives by 1,000 to ease interpretation.

Table 4: Elasticities of labor supply

	Based on dep. var.: Will work in this scenario		Based on dep. var.: Prefers this scenario over other	
	No social preferences (1)	With social preferences (2)	No social preferences (3)	With social preferences (4)
Panel A: Marshallian elasticity of labor supply with respect to:				
Wage	-0.12***	-0.10***	-0.11***	-0.097***
	(0.012)	(0.012)	(0.016)	(0.016)
Net-of-tax rate	-0.12***	-0.38***	-0.11***	-0.33***
	(0.012)	(0.023)	(0.016)	(0.022)
Wage-tax elasticity wedge	0	0.28***	0	0.23***
	(.)	(0.024)	(.)	(0.019)
Panel B: Hicksian elasticity of labor supply with respect to:				
Wage	0.42***	0.47***	0.50***	0.53***
	(0.019)	(0.021)	(0.021)	(0.021)
Net-of-tax rate	0.42***	-0.24***	0.50***	-0.0097
	(0.019)	(0.054)	(0.021)	(0.046)
Wage-tax elasticity wedge	0	0.71***	0	0.54***
	(.)	(0.060)	(.)	(0.049)
Panel C: Frisch elasticity of labor supply with respect to:				
Wage	1.31***	1.22***	1.86***	1.78***
	(0.11)	(0.10)	(0.15)	(0.15)
Net-of-tax rate	1.31***	0.63***	1.86***	1.17***
	(0.11)	(0.080)	(0.15)	(0.11)
Wage-tax elasticity wedge	0	0.59***	0	0.62***
	(.)	(0.065)	(.)	(0.063)
Respondents	4,671	4,671	5,381	5,381

Notes: Each cell shows the average simulated elasticity or difference in elasticities based on the model specified in the column header. Elasticities and differences are simulated using the parameters of the model, at the reported (for non-tax variables) or tax-simulated values of each respondent (one for each respondent), and then averaged over respondents. In each panel, the wage-tax elasticity wedge is the wage elasticity less the net-of-tax rate elasticity. Standard errors computed by delta method in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5: Robustness of the wage-tax elasticity wedge

		Frisch wage-tax elasticity wedge, estimated based on:	
		Will work in this scenario (1)	Prefers this scenario over other (2)
0	Main estimates	0.59***	0.62***
		(0.065)	(0.063)
1	Fifth order polynomial in G	0.47***	0.54
		(0.086)	(1.25)
2	G interacted with $\log(c)$ and $\log(\bar{L}-h)$	0.25***	0.32**
		(0.097)	(0.13)
3	Quadratic utility	0.61***	0.65***
		(0.079)	(0.090)
4	β_c and β_h varies with individual characteristics	0.56***	0.59***
		(0.063)	(0.061)
5	Using non-imputed consumption	0.76***	0.75***
		(0.11)	(0.10)
6	Consumption imputed based on Empirical Bayes	0.42***	0.40***
		(0.070)	(0.071)
7	No $\log(assets_i + spouseincome_i + savings_{itj})$ controls	1.01***	1.03***
		(0.092)	(0.080)
8	Exclude pairs since observation of the intransitivity prompt	0.46***	0.53***
		(0.057)	(0.059)
9	Exclude respondents who ever made an intransitive choice	0.51***	0.54***
		(0.064)	(0.062)
10	Exclude inattentive or impatient or speeding respondents	0.42***	0.43***
		(0.059)	(0.052)
11	Median estimate	0.57***	0.56***
		(0.065)	(0.068)
12	Reweight for demographics	0.52***	0.65***
		(0.10)	(0.13)

Notes: Standard errors computed by delta method in parentheses, unless stated otherwise below. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Each cell shows the average simulated difference in Frisch elasticities (wage elasticity minus net-of-tax rate elasticity) based on the dependent variable in the column header. Each row shows a specification deviation from our main estimates in Row 0. Row 1: The utility function uses a fifth-order polynomial in the tax-funded government expenditure. Row 2: The utility function interacts tax-funded government expenditure with log consumption and leisure. Row 3: The utility function uses a quadratic utility specification instead of a translog specification. Row 4: The utility function models heterogeneity in the main coefficients of consumption and leisure. Specifically, these two coefficients are allowed to vary linearly with sex, marital status (married or not), sex interacted with marital status, a quadratic in log age, a quadratic in the number of children in the family (top-coded at 9), an indicator for education level of high-school graduate or less, and an indicator for education level of four-year college degree or more. Row 5: We restrict the sample to pairs 5 to 8 and use the (non-imputed) consumption information of each scenario directly. Row 6: We use empirical Bayes prediction as an alternative imputation method for consumption. Specifically, we allow the coefficient on log take-home earnings to be random, and add (non-random) interactions of the log take-home earnings with each of sex, marital status, sex interacted with marital status, an indicator for high-school graduate or less, an indicator for four-year college degree or more, a quadratic in log age, and a quadratic in the number of children in the family. Row 7: Our main estimates control for a quadratic in the log of the sum of assets, spousal income, and the scenario's savings; we exclude these two terms in this row. Row 8: Our survey shows text below the question grid that informs respondents about possible intransitivity in choices when committed; this row uses only pairs before observation of the text. Row 9: We exclude respondents who ever made an intransitive choice. This excludes respondents who made the choice and then changed their minds later as well. Row 10: We exclude respondents who were inattentive in our two attention check questions, clicked "next" more than twenty times when viewing the instruction slides, or answered any of the overall, vignette, and government expenditure liking questions in less than half the median time. Row 11: We report the median simulated wage-tax elasticity wedge instead of the mean. Standard errors in this row are bootstrapped standard errors based on 199 replications clustered at the respondent level, with each replication estimating the model and computing the simulated median. Row 12: We weight the estimation and simulated differences using post-stratification weights constructed based on the variables in Table 1.

Table 6: Heterogeneity in the wage-tax elasticity wedge by opinion of government

	Wage-tax elasticity wedge, differenced across specified heterogeneity variable		
	Marshallian (1)	Hicksian (2)	Frisch (3)
General opinion of government	0.029	0.041	0.26***
	(0.020)	(0.044)	(0.083)
Importance of government	-0.011	-0.037	0.020
	(0.021)	(0.055)	(0.059)
Trust in government	0.044	0.18***	0.26**
	(0.028)	(0.067)	(0.13)
Programs benefit people like me	0.044*	0.15***	0.15*
	(0.023)	(0.057)	(0.082)
Government revenue allocation is fair	0.034*	0.046	0.20***
	(0.018)	(0.040)	(0.064)

Notes: Standard errors computed by delta method in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Sample and estimation method follows that of Tables 3 and Tables 4, see Section 3.4 for details. Each cell reports the Frisch wage elasticity less the Frisch net-of-tax rate elasticity, difference across the heterogeneity variable specified in the row title. Each heterogeneity variable is constructed by taking the average standardized score of two questions, and then dichotomized such that a value of 1 indicates a good opinion while a 0 indicates a bad opinion.

Table 7: Relationship between liking an expenditure category and the wage-tax elasticity wedge

	Average partial effect of liking the program on the wedge specified (1)
Marshallian wage-tax elasticity wedge	0.27**
	(0.13)
Hicksian wage-tax elasticity wedge	0.55**
	(0.27)
Frisch wage-tax elasticity wedge	0.69*
	(0.41)
Respondents	3,136

Notes: Each row shows the average partial effect of liking a program on the specified wage-tax elasticity wedge, where the net-of-tax rate elasticity is specific to the program. Standard errors computed by delta method in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Estimation method follows that of Tables 3 and Tables 4, see Section 4.3 for details. The underlying sample is that from the same tables, excluding respondents who failed either of two inattention check questions, clicked "next slide" more than 20 times in the instructions slides, or answered any of the overall, vignette, and government expenditure liking questions in less than half the median time taken by all respondents.

Table 8: The relationship between government-related social preferences and the elasticity of taxable income

	Dependent variable: ETI. Government-related social preferences proxy is:						
	Confidence in government	Confidence in political parties	Confidence in parliament	Confidence in civil service	Income should be made more equal	Gvt should increase ownership of businesses	Proud to be a citizen
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main estimates							
Gvt-related social preferences proxy	-0.50***	-0.36***	-0.41***	-0.49*	-0.047	-0.16**	-0.37***
	(0.079)	(0.073)	(0.12)	(0.25)	(0.057)	(0.078)	(0.13)
Observations	1,701	1,701	1,701	1,701	1,701	1,701	1,701
Number of studies	60	60	60	60	60	60	60
Panel B: Robustness							
1 Before-deduction elasticities only	-0.36***	-0.28	-0.23*	-0.19	-0.044	-0.00071	-0.28***
	(0.12)	(0.17)	(0.12)	(0.11)	(0.046)	(0.053)	(0.097)
2 After-deduction elasticities only	-0.53***	-0.38***	-0.46***	-0.76**	-0.051	-0.30**	-0.48*
	(0.082)	(0.074)	(0.13)	(0.36)	(0.12)	(0.11)	(0.25)
3 Include Neisser's contextual factors	-0.44***	-0.48***	-0.48***	-0.73***	-0.12	-0.21***	-0.010
	(0.078)	(0.091)	(0.072)	(0.21)	(0.094)	(0.060)	(0.12)
4 Include country-group FE	-0.51***	-0.36***	-0.41***	-0.42	-0.027	-0.24***	-0.28*
	(0.089)	(0.082)	(0.15)	(0.29)	(0.061)	(0.083)	(0.15)
5 Include country-group and time FE	-0.54***	-0.42***	-0.43***	-0.68*	-0.042	-0.31***	-0.29
	(0.080)	(0.064)	(0.14)	(0.38)	(0.10)	(0.11)	(0.20)
6 Control for labor share subsidy	-0.54***	-0.42***	-0.50***	-0.56**	0.012	-0.14	-0.38**
	(0.092)	(0.073)	(0.12)	(0.25)	(0.038)	(0.088)	(0.14)

Notes: Standard errors clustered by study in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises all ETI estimates in Neisser (2021), excluding one study from Israel. The government-related social preferences proxies are country-wave level averages from the WVS and the EVS (Inglehart et al. 2022; EVS 2022); we match observations to the nearest WVS or EVS year using the mean study year. All estimates control for the regression technique used (five possible categories, including the omitted base category), whether and how the study controls for income (five categories), the difference length (e.g. 1 year) for the first-differences technique used in estimation (4 categories), whether the study weights by income, whether the study restricts the sample by age, the income restriction used by the study if any (five categories), the publication decade (3 categories), and the decade that the tax policy changed (three categories). See Neisser's Table 2 and 3 column 6 for further details on the above variables. All estimates additionally control for whether the elasticity is a before-deductions (BD) or after-deductions (AD) elasticity, and the interactions of this dummy variable with all control variables listed above. Panel A shows the main estimates; Panel B shows robustness estimates that deviate in the specified manner. Rows 1 and 2: The sample is restricted to BD and AD elasticities respectively, and we no longer control for the BD-or-AD dummy and its interactions. Row 3: We additionally control for all contextual factors except the top 1% income share from Neisser's Table 5 and their interactions with the BD-or-AD dummy. Row 4: We include the country-group FEs and their interactions with BD-or-AD dummy. Row 5: We include the fixed effects from Row 4, and additionally include FEs for the year ranges 1981–1984, 1989–1993, 1994–1998, 1999-2004, 2005–2009, 2010–2014, 2017–2022 (and their interactions with the BD-or-AD dummy). These years are WVS survey years, and the five EVS surveys match (uniquely) to the first, second, fourth, fifth, and seventh year ranges. Row 6: We additionally control for the share of public spending on the provision of child care, preschool, and elderly care, available in Kleven (2014), and it's interaction with BD-or-AD dummy.

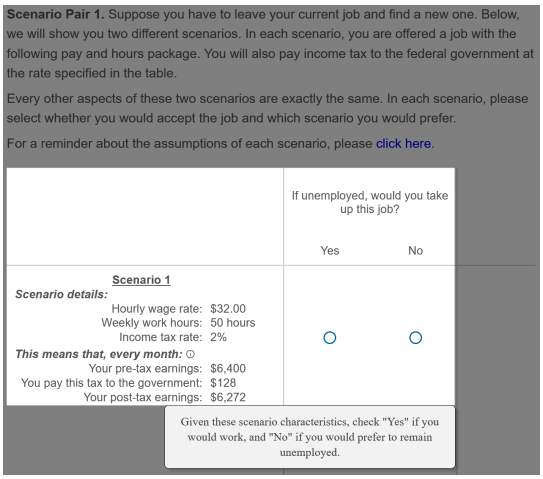
Online Appendix for
Wages, taxes, and labor supply elasticities: The role of social preferences

By Janjala Chirakijja and Pinchuan Ong

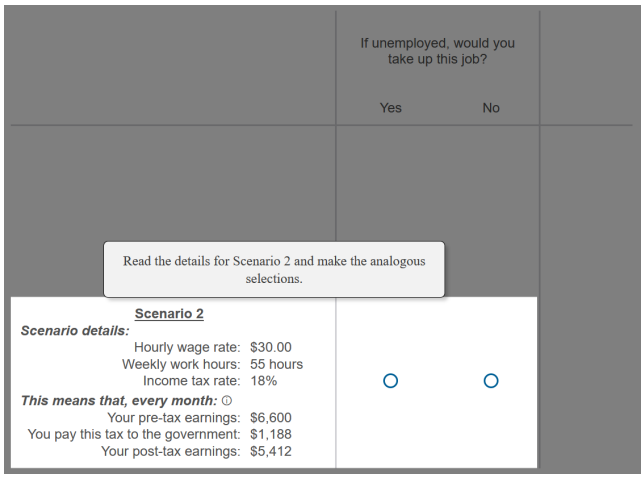
A Appendix tables and figures

Appendix Figure A1: Example of scenarios presentation for pairs 1 to 4

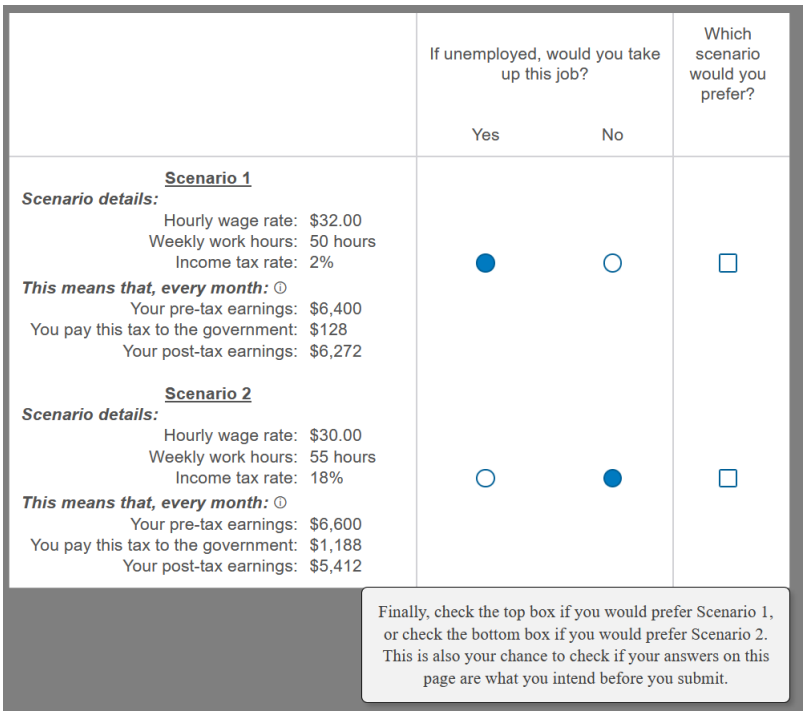
(a) Work or not in first scenario



(b) Work or not in second scenario



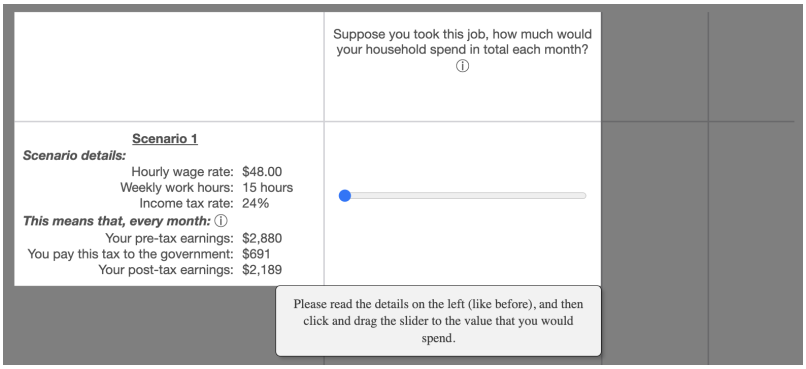
(c) Choice between scenarios



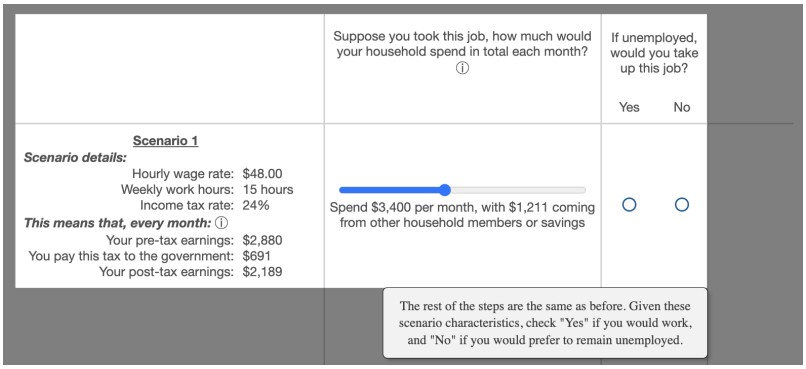
Notes: The figure shows an example of the scenario presentation for pairs 1 to 4. Panels A and B shows that the other scenario of the pair was hidden when the respondent was asked to make a choice of whether to work or not. Panel C shows that both scenarios were shown again when the respondent was asked to choose between the two. The guided website tour (white focus on black background, with explanatory text) was shown only in pair 1 of the first four pairs.

Appendix Figure A2: Example of scenarios presentation for pairs 5 to 8

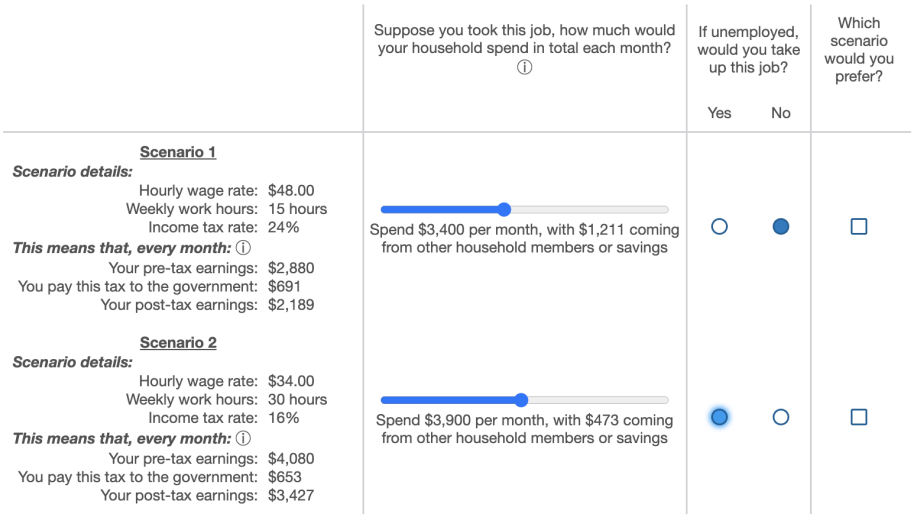
(a) Consumption amount



(b) Work or not



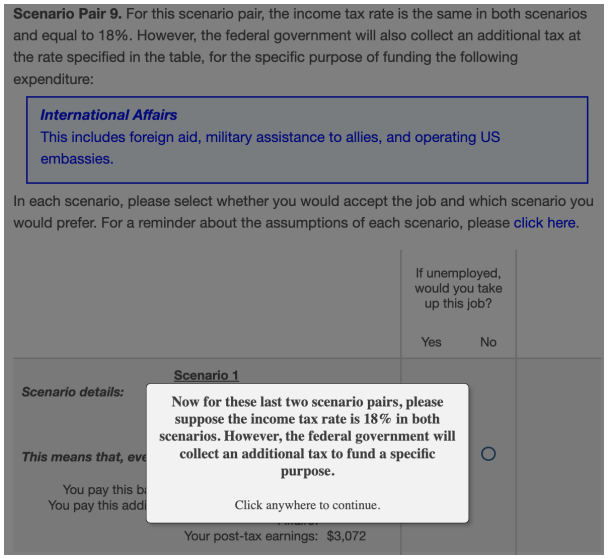
(c) Choice between scenarios



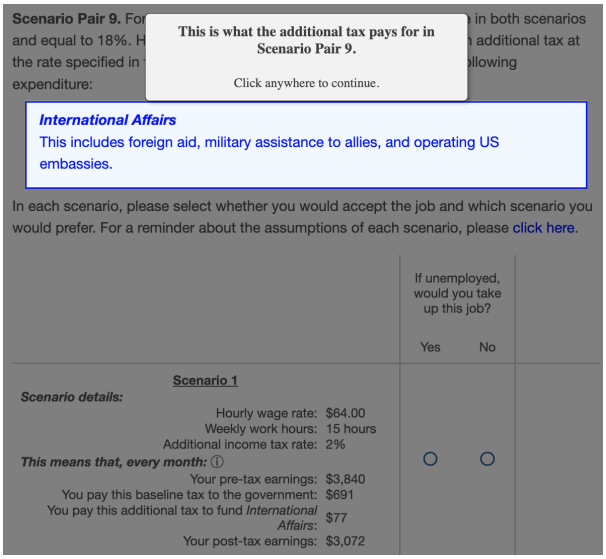
Notes: The figure shows an example of the scenario presentation for pairs 5 to 8, omitting a repeat of the instructions shown above the questionnaire matrix. Panels A and B show that the choice of whether to work was hidden when eliciting the amount of consumption using the slider bar shown, and revealed only after the respondent made the consumption choice. The second scenario was hidden when the respondent made choices for the first, and vice versa, analogous to Appendix Figure A1. Panel C shows the full layout, revealed only when the respondent was asked to choose between the two scenarios. The guided website tour (white focus on black background, with explanatory text) was shown only in pair 5 of pairs 5 to 8.

Appendix Figure A3: Example of scenarios presentation for pairs 9 and 10

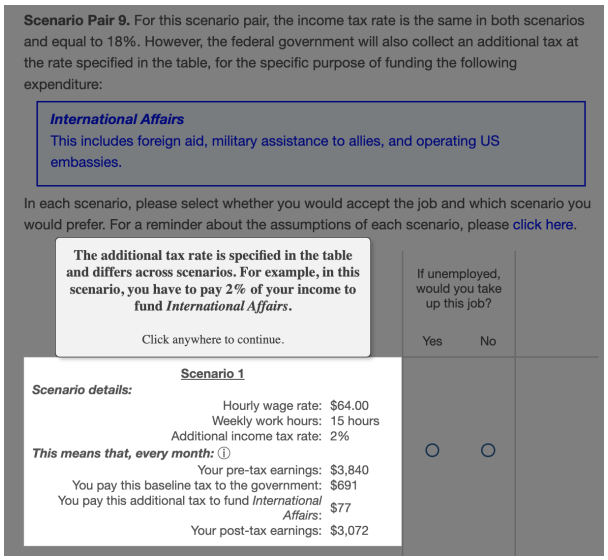
(a) Screen on entering the ninth pair



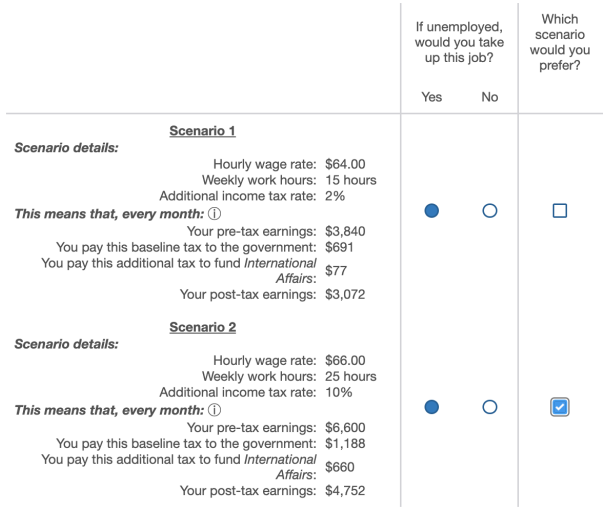
(b) First transition of website tour



(c) Second transition of website tour

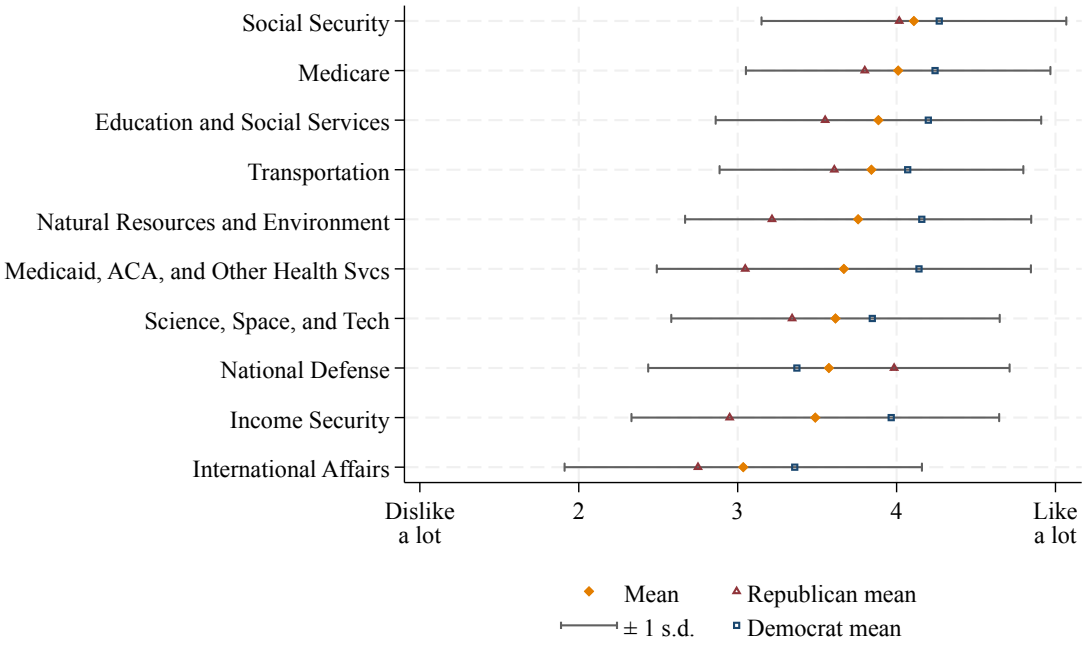


(d) Full matrix



Notes: The figure shows an example of the scenario presentation for pairs 9 and 10. Panels A shows the instructions displayed on entering the ninth pair. Panels B and C similarly show the information that we highlighted to respondents on the first two transitions of the guided website tour. Panel D shows the full layout. The second scenario was hidden when the respondent made choices for the first, and vice versa, analogous to Appendix Figure A1. The transition in Panel B was also displayed in pair 10 to highlight the new program that the change in program being funded.

Appendix Figure A4: How much respondents like government programs



Notes: The figure shows means and standard deviations of responses to the question "To what extent do you like or dislike that your tax money is used to fund each [federal government spending] category?" The sample comprises respondents with valid in-vignette consumption data, excluding respondents in the top and bottom 5 percent of actual work hours, consumption, and simulated tax rate. Each respondent received five of the ten categories shown. Each category was displayed with a tooltip which, on mouse hover, would provide more details. Republican (Democrat) mean indicates the mean computed among respondents who reported a Republican (Democrat) political affiliation.

Appendix Table A1: Horse-race regression of consumption and earnings

	Log(consumption) (1)
Log(Take-home earnings)	0.37***
	(0.036)
Log(Pre-tax earnings)	0.024
	(0.036)
Observations	43502
Respondents	5440

Notes: Standard errors clustered by respondents in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises respondent-pair-scenarios with valid in-vignette consumption data, excluding respondents in the top and bottom 5 percent of actual work hours, consumption, and simulated tax rate. The specification controls for individual, pair, and scenario order fixed effects.

Appendix Table A2: Frisch elasticities of labor supply, by sex and marital status

	Subsample:			
	Non-married men (1)	Non-married women (2)	Married men (3)	Married women (4)
Panel A: Dependent variable: Will work in this scenario				
Frisch wage elasticity	1.35***	1.24***	1.03***	0.87***
	(0.19)	(0.17)	(0.22)	(0.13)
Frisch net-of-tax rate elasticity	0.69***	0.60***	0.58***	0.36***
	(0.18)	(0.16)	(0.22)	(0.13)
Frisch wage-tax elasticity wedge	0.66***	0.64***	0.45***	0.52***
	(0.072)	(0.066)	(0.075)	(0.062)
Respondents in subsample	1,136	1,056	1,244	1,235
Panel B: Dependent variable: Prefers this scenario over other				
Frisch wage elasticity	1.76***	1.82***	1.29***	1.76***
	(0.20)	(0.20)	(0.24)	(0.20)
Frisch net-of-tax rate elasticity	1.14***	1.19***	0.75***	1.14***
	(0.19)	(0.19)	(0.23)	(0.19)
Frisch wage-tax elasticity wedge	0.62***	0.63***	0.54***	0.62***
	(0.070)	(0.064)	(0.084)	(0.070)
Respondents in subsample	1,304	1,251	1,457	1,304

Notes: Each cell shows the average simulated elasticity or wage-tax elasticity wedge in elasticities for the subsample specified in the column header. To compute the elasticities and wedge, we first estimate the extended model from Table 5 row 4, and then simulate the elasticities and wedges for the indicated subsample at the reported (for non-tax variables) or tax-simulated values of each respondent (one for each respondent), and then averaged over respondents. Standard errors computed by delta method in parentheses. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

B Derivations

B.1 Incorporating tax contribution of others

In this section, we elaborate the utility function to explicitly include social preferences from the tax contribution of other people. We show that the model simplifies to our main model used in Section 2 given an assumption that is likely to hold in our vignette experiment.

In any time period, a person with initial assets level a_i chooses consumption c_i, next period assets a_i', and labor supply h_i to solve the utility maximization problem

$$V\left(a_i, w_i, \tau, N_i\right) = \max_{c_i, a_i', h_i} \tilde{U}\left(c_i, h_i, g_i, g_{-i}, G\right) + \beta E\left[V\left(a_i', w_i', \tau', N_i'\right)\right] \tag{B1}$$

$$\text{s.t. } c_i + \frac{1}{1+r} a_i' = a_i + (1-\tau) w_i h_i + N_i, \tag{B2}$$

$$g_i = \tau w_i h_i, \tag{B3}$$

$$g_{-i} = \tau Y_{-i}, \tag{B4}$$

$$G = G_0 + \tau Y_{-i} + g_i, \tag{B5}$$

where w_i is the person's wage, τ is the tax rate on earnings from work $w_i h_i$, N_i is non-labor income, β is an impatience parameter, r is the interest rate, $V(\cdot)$ is the indirect utility function, and $E[\cdot]$ is the expectations operator. Equation (B2) is a usual budget constraint which specifies that consumption c_i and savings $\frac{1}{1+r} a_i' - a_i$ equals post-tax earnings and non-labor income.

Compared to the usual labor supply model, per-period utility $\tilde{U}(\cdot)$ in Equation (B1) depends on three additional factors: the individual's own tax contribution g_i (reflecting warm glow from own contribution), the other people's tax contribution g_{-i} (reflecting warm glow from others' contribution), and the total government expenditure G (reflecting pure altruism, which by definition would not distinguish between a contribution by self or by others). g_i is equivalent to the tax rate applied to the individual's total earnings from work by Equation (B3), and g_{-i} is equivalent to the same income tax rate applied to the total earnings of everyone else Y_{-i} by Equation (B4). Equation (B5) computes total government expenditure as the sum of non-income tax revenue G_0, g_i and g_{-i}. Notice that $\tilde{U}(\cdot)$ depends (through g_{-i} and G) on the common tax rate τ. Hence, when tax rates increase, the individual gets social preferences utility not just from her own contribution but also the increased contributions of others. This modeling aspect reflects the fact that tax changes in practice often apply to broad groups of people, while wage changes are often perceived by the individual to be specific to herself.

While conceptually easy to describe, the maximization problem (B1) does not yield equations that can be easily estimated. This is because estimating the problem directly requires data on the individual's expectations of Y_{-i}, which is difficult to elicit. However, notice that substituting Equation B4 and (B5) into the utility function yields $\tilde{U}\left(c, h, g_i, \tau Y_{-i}, G_0 + g_i + \tau Y_{-i}\right)$, with G_0 and Y_{-i} exogenous to the individual. Hence, treating these as parameters in the utility function, we can rewrite utility as

$$U\left(c_i, h_i, g_i, \tau\right) = \tilde{U}\left(c, h, g_i, \tau Y_{-i}, G_0 + g_i + \tau Y_{-i}\right). \tag{B6}$$

The maximization problem then becomes

$$V\left(a_i, w_i, \tau, N_i\right) = \max_{c_i, a_i', h_i} U\left(c_i, h_i, g_i, \tau\right) + \beta E\left[V\left(a_i', w_i', \tau', N_i'\right)\right] \tag{B7}$$

subject to Equations (B2) and (B3). This formulation yields estimating equations involving τ, which are more easily observed in the data.

For our empirical analysis, we specialize this utility function to

$$U\left(c_i, h_i, g_i, \tau\right) = u\left(c_i, h_i\right) + v\left(g_i, \tau\right) \tag{B8}$$

so that we have an easily interpretable social preferences term $v\left(g_i, \tau\right)$. Since taxes paid are highly correlated with tax rates, we expect the model misspecification that comes from ignoring the dependence of $v\left(g_i, \tau\right)$ on τ to be relatively benign. In that case, we can further simplify to

$$U\left(c_i, h_i, g_i, \tau\right) = u\left(c_i, h_i\right) + \tilde{v}\left(g_i\right). \tag{B9}$$

This is equivalent to the utility function that we use in our main analysis, as described in Section 2. The first derivative of $\tilde{v}\left(g_i\right)$ will be positive if people had positive social preferences—broadly defined, since social preferences in our model also includes the tax money that comes from other individuals g_{-i} (in addition to the usual discussion about pure altruism and warm glow from the individual's own contribution).

B.2 Main model derivation

The utility maximization problem from Section 2, slightly rewritten, is

$$V\left(a_i, w_i, \tau, N_i\right) = \max_{c_i, a_i', h_i} U\left(c_i, h_i, g_i\right) + \beta E\left[V\left(a_i', w_i', \tau', N_i'\right)\right] \tag{B10}$$

$$\text{s.t. } c_i = A_i + \left(1 - \tau\right) w_i h_i + N_i, \tag{B11}$$

$$A_i = a_i - \frac{1}{1+r} a_i', \tag{B12}$$

$$g_i = \tau w_i h_i, \tag{B13}$$

where utility $U(\cdot)$ depends on consumption c_i, labor supply h_i, and tax-funded government expenditure g_i, the state variables for the maximization problem are asset level a_i, wage w_i, tax rate on earnings τ, and non-labor income N_i, β is an impatience parameter, r is the interest rate, $V(\cdot)$ is the indirect utility function, $E[\cdot]$ is the expectations operator, and the prime symbol ($'$) denotes the next period. A_i is the intertemporal resource allocation amount, useful for connecting this to static models.

Let λ_i be the Lagrangian multiplier on Equation (B11). The (intratemporal) first order conditions for c_i and h_i are, respectively,

$$U_c(c_i, h_i, g_i) = \lambda_i \tag{B14}$$

and

$$U_h(c_i, h_i, g_i) + \tau w_i U_g(c_i, h_i, g_i) + \lambda_i (1-\tau) w_i = 0. \tag{B15}$$

Substituting Equation (B14) into Equation (B15) and rewriting in matrix form (more convenient for derivation of the Marshallian and Hicksian elasticities), we have

$$\mathcal{M}(h_i; c_i, A_i, a_i, w_i, \tau, N_i) \equiv DU(c_i, h_i, g_i) \begin{bmatrix} (1-\tau) w_i \\ 1 \\ \tau w_i \end{bmatrix} = 0, \tag{B16}$$

where D is the directional derivative operator so that

$$DU(c_i, h_i, g_i) \equiv \begin{bmatrix} U_c(c_i, h_i, g_i) & U_h(c_i, h_i, g_i) & U_g(c_i, h_i, g_i) \end{bmatrix}$$

is the Jacobian matrix of $U(c_i, h_i, g_i)$, and the optimal-choice c_i and g_i satisfy Equations (B11) and (B13) respectively.

B.2.1 Marshallian elasticities

In a life-cycle framework, the Marshallian elasticity holds constant the intertemporal allocation of resources A_i (see Blundell and MaCurdy 1999, Section 4.2.2). Implicit differentiation of Equation (B16) gives

$$\epsilon_w^M \equiv \left. \frac{\partial \log h_i}{\partial \log w_i} \right|_{A_i} = -\frac{w_i \frac{\partial \mathcal{M}(h_i)}{\partial w_i}}{h_i \frac{\partial \mathcal{M}(h_i)}{\partial h_i}} \tag{B17}$$

and

$$\epsilon_{1-\tau}^{M} \equiv \left.\frac{\partial \log h_i}{\partial \log (1-\tau)}\right|_{A_i} = -\frac{(1-\tau)\frac{\partial \mathcal{M}(h_i)}{\partial (1-\tau)}}{h\frac{\partial \mathcal{M}(h_i)}{\partial h_i}}, \tag{B18}$$

where we have suppressed the dependence of the elasticities on i to reduce notational burden. Similarly suppressing the dependence of derivatives $U(c_i, h_i, g_i)$ on their arguments and on i, the derivatives of the $\mathcal{M}(h_i)$ terms in Equations (B17) and (B18) are

$$\frac{\partial \mathcal{M}(h_i)}{\partial h_i} = \begin{bmatrix} (1-\tau)w_i & 1 & \tau w_i \end{bmatrix} D^2U \begin{bmatrix} (1-\tau)w_i \\ 1 \\ \tau w_i \end{bmatrix},$$

$$\frac{\partial \mathcal{M}(h_i)}{\partial w_i} = \begin{bmatrix} (1-\tau)h_i & 0 & \tau h_i \end{bmatrix} D^2U \begin{bmatrix} (1-\tau)w_i \\ 1 \\ \tau w_i \end{bmatrix} + DU \begin{bmatrix} (1-\tau) \\ 0 \\ \tau \end{bmatrix},$$

and

$$\frac{\partial \mathcal{M}(h_i)}{\partial (1-\tau)} = \begin{bmatrix} w_i h_i & 0 & -w_i h_i \end{bmatrix} D^2U \begin{bmatrix} (1-\tau)w_i \\ 1 \\ \tau w_i \end{bmatrix} + DU \begin{bmatrix} w_i \\ 0 \\ -w_i \end{bmatrix},$$

where D^2U is the Hessian matrix of $U(c_i, h_i, g_i)$. Specializing to the separable social preferences utility in Equation (4) and evaluating (we use the Python symbolic mathematics package SymPy), Equations (B17) and (B18) simplify to

$$\epsilon_{w}^{M} = \frac{-w_i(1-\tau)\left\{\left[U_c + \frac{\tau}{1-\tau}(U_g + U_{gg}g_i)\right] + h_i\left[(1-\tau)w_i U_{cc} + U_{ch}\right]\right\}}{h_i\left[(1-\tau)^2 w_i^2 U_{cc} + 2(1-\tau)w_i U_{ch} + U_{hh} + \tau^2 w_i^2 U_{gg}\right]} \tag{B19}$$

and

$$\epsilon_{1-\tau}^{M} = \frac{-w_i(1-\tau)\left\{\left[U_c - (U_g + U_{gg}g_i)\right] + h_i\left[(1-\tau)w_i U_{cc} + U_{ch}\right]\right\}}{h\left[(1-\tau)^2 w_i^2 U_{cc} + 2(1-\tau)w_i U_{ch} + U_{hh} + \tau^2 w_i^2 U_{gg}\right]} \tag{B20}$$

respectively.

B.2.2 Hicksian elasticities

The expenditure minimization problem is

$$e\left(a_i, w_i, \tau, \bar{U}_i\right) = \min_{c_i, a_i', h_i} c_i - A_i - (1-\tau)w_i h_i \tag{B21}$$

$$\text{s.t. } U(c_i, h_i, g_i) = \bar{U}_i \tag{B22}$$

and Equations (B12) and (B13). Let $h^H\left(a_i, w_i, \tau, \bar{U}_i\right)$ denote the solution. Duality gives

$$h^H\left(a_i, w_i, \tau, \bar{U}_i\right) = h\left(a_i, w_i, \tau, e\left(a_i, w_i, \tau, \bar{U}_i\right)\right), \tag{B23}$$

where $h(\cdot)$ is the Marshallian labor supply function, so that holding A_i constant and differentiating with respect to w_i and $(1-\tau)$ gives, respectively,

$$\begin{aligned}\left.\frac{\partial h_i^H}{\partial w_i}\right|_{A_i,\bar{U}_i} &= \left.\frac{\partial h_i}{\partial w_i}\right|_{A_i} + \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left.\frac{\partial e\left(a_i, w_i, \tau, \bar{U}_i\right)}{\partial w_i}\right|_{A_i,\bar{U}_i} \\ &= \left.\frac{\partial h_i}{\partial w_i}\right|_{A_i} - (1-\tau)\, h_i \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left(1 + \frac{\tau U_g}{(1-\tau)\, U_c}\right)\end{aligned} \tag{B24}$$

and

$$\begin{aligned}\left.\frac{\partial h_i^H}{\partial (1-\tau)}\right|_{A_i,\bar{U}_i} &= \left.\frac{\partial h_i}{\partial (1-\tau)}\right|_{A_i} + \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left.\frac{\partial e\left(a_i, w_i, \tau, \bar{U}_i\right)}{\partial (1-\tau)}\right|_{A_i,\bar{U}_i} \\ &= \left.\frac{\partial h_i}{\partial (1-\tau)}\right|_{A_i} - w_i h_i \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left(1 - \frac{U_g}{U_c}\right),\end{aligned} \tag{B25}$$

where we have used the envelope theorem and the first order condition with respect to c_i in the second equality. The elasticities are thus

$$\epsilon_w^H \equiv \left.\frac{\partial \log h_i^H}{\partial \log w_i}\right|_{A_i,\bar{U}_i} = \epsilon_w^M - (1-\tau)\, w_i \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left(1 + \frac{\tau U_g}{(1-\tau)\, U_c}\right) \tag{B26}$$

and

$$\epsilon_{1-\tau}^H \equiv \left.\frac{\partial \log h_i^H}{\partial \log (1-\tau)}\right|_{A_i,\bar{U}_i} = \epsilon_{1-\tau}^M - (1-\tau)\, w_i \left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} \left(1 - \frac{U_g}{U_c}\right), \tag{B27}$$

with $\left.\frac{\partial h_i}{\partial N_i}\right|_{A_i}$ obtained by implicit differentiation of Equation (B16):

$$\left.\frac{\partial h_i}{\partial N_i}\right|_{A_i} = -\frac{\frac{\partial \mathcal{M}(h_i)}{\partial N_i}}{\frac{\partial \mathcal{M}(h_i)}{\partial h_i}} = -\frac{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} D^2 U \begin{bmatrix} (1-\tau)\, w_i \\ 1 \\ \tau w_i \end{bmatrix}}{\frac{\partial \mathcal{M}(h_i)}{\partial h_i}}. \tag{B28}$$

Substituting in the expressions from Equations (B17) and (B18), and simplifying, we get

$$\epsilon_w^H = \frac{\text{Num}_w^H}{h_i \frac{\partial \mathcal{M}(h_i)}{\partial h_i}}, \tag{B29}$$

$$\text{Num}_w^H = \left[\begin{array}{ccc} \frac{U_g}{U_c}\tau w_i h_i & 0 & -\tau w_i h_i \end{array} \right] D^2U \left[\begin{array}{c} (1-\tau)\, w_i \\ 1 \\ \tau w_i \end{array} \right] + DU \left[\begin{array}{c} -(1-\tau)\, w_i \\ 0 \\ -\tau w_i \end{array} \right],$$

and

$$\epsilon_{1-\tau}^H = \frac{\text{Num}_{1-\tau}^H}{h_i \frac{\partial \mathcal{M}(h_i)}{\partial h_i}}, \tag{B30}$$

$$\text{Num}_{1-\tau}^H = \left[\begin{array}{ccc} -\frac{U_g}{U_c}(1-\tau)\, w_i h_i & 0 & (1-\tau)\, w_i h_i \end{array} \right] D^2U \left[\begin{array}{c} (1-\tau)\, w_i \\ 1 \\ \tau w_i \end{array} \right] + DU \left[\begin{array}{c} -(1-\tau)\, w_i \\ 0 \\ (1-\tau)\, w_i \end{array} \right].$$

Finally, specializing to the separable social preferences utility in Equation (4) gives

$$\epsilon_w^H = \frac{-w_i (1-\tau) \left\{ \left[U_c + \frac{\tau}{1-\tau} \left(U_g + U_{gg} g_i \right) \right] - \frac{\tau}{1-\tau} h_i \frac{U_g}{U_c} \left[(1-\tau)\, w_i U_{cc} + U_{ch} \right] \right\}}{h_i \left[(1-\tau)^2 w_i^2 U_{cc} + 2(1-\tau)\, w_i U_{ch} + U_{hh} + \tau^2 w_i^2 U_{gg} \right]} \tag{B31}$$

and

$$\epsilon_{1-\tau}^H = \frac{-w_i (1-\tau) \left\{ \left[U_c - \left(U_g + U_{gg} g_i \right) \right] + h_i \frac{U_g}{U_c} \left[(1-\tau)\, w_i U_{cc} + U_{ch} \right] \right\}}{h_i \left[(1-\tau)^2 w_i^2 U_{cc} + 2(1-\tau)\, w_i U_{ch} + U_{hh} + \tau^2 w_i^2 U_{gg} \right]}. \tag{B32}$$

Somewhat surprising, the income effect term $h_i \left[(1-\tau)\, w_i U_{cc} + U_{ch} \right]$ does not disappear from the compensated elasticity. (We ignore the denominator in this discussion for brevity.) This occurs because tax-funded government expenditure is outside the control of the individual (an agency problem), and hence does not envelope out when holding utility constant. To understand this, consider the extreme case of $U_g = U_c$ (i.e. when the person derives as much marginal utility from tax-funded government expenditure as from consumption). In this case, the Hicksian net-of-tax rate elasticity in Equation (B32) is the same as the Marshallian counterpart in Equation (B20)—no "Hicksian compensation" is needed to keep utility constant because government expenditures automatically compensate for the income effect. (Valuing consumption and taxation equally means that there is no longer an agency problem.) On the other hand, the Hicksian wage elasticity in Equation (B31) is larger than the Marshallian counterpart in Equation (B19)—on top of compensating for the income effect, we have to add a $-\frac{\tau}{1-\tau} h_i \left[(1-\tau)\, w_i U_{cc} + U_{ch} \right]$ term to compensate for the difference

induced by the agency problem.

B.2.3 Frisch elasticities

The Frisch elasticities hold constant the marginal utility of wealth λ_i. To do this, rewrite Equations (B14) and (B15) in matrix form to get

$$\boldsymbol{\mathcal{F}}(c_i, h_i) \equiv \begin{bmatrix} U_c(c_i, h_i, g_i) \\ U_h(c_i, h_i, g_i) + \tau w_i U_g(c_i, h_i, g_i) \end{bmatrix} + \begin{bmatrix} -\lambda_i \\ \lambda_i (1-\tau) w_i \end{bmatrix} = \mathbf{0}.$$

This is a system of two equations in c_i and h, with parameters τ, w_i, and λ_i. (g_i is not a parameter since it is pinned down by Equation (B13)). Implicit differentiation gives

$$\begin{bmatrix} \frac{\partial c_i}{\partial w_i}\Big|_{\lambda_i} \\ \frac{\partial h_i}{\partial w_i}\Big|_{\lambda_i} \end{bmatrix} = -\left(D\boldsymbol{\mathcal{F}}(c_i, h_i)\right)^{-1} \frac{\partial \boldsymbol{\mathcal{F}}(c_i, h_i)}{\partial w_i}\Big|_{\lambda_i}, \tag{B33}$$

and

$$\begin{bmatrix} \frac{\partial c_i}{\partial (1-\tau)}\Big|_{\lambda_i} \\ \frac{\partial h_i}{\partial (1-\tau)}\Big|_{\lambda_i} \end{bmatrix} = -\left(D\boldsymbol{\mathcal{F}}(c_i, h_i)\right)^{-1} \frac{\partial \boldsymbol{\mathcal{F}}(c_i, h_i)}{\partial (1-\tau)}\Big|_{\lambda_i}, \tag{B34}$$

where

$$\frac{\partial \boldsymbol{\mathcal{F}}(c_i, h_i)}{\partial w_i}\Big|_{\lambda_i} = \begin{bmatrix} \tau h_i U_{cg} \\ \tau h_i U_{hg} + \tau U_g + \tau^2 w_i h_i U_{gg} + (1-\tau)\lambda_i \end{bmatrix},$$

$$\frac{\partial \boldsymbol{\mathcal{F}}(c_i, h_i)}{\partial (1-\tau)}\Big|_{\lambda_i} = \begin{bmatrix} -w_i h_i U_{cg} \\ -w_i h_i U_{hg} - w_i U_g - \tau w_i^2 h_i U_{gg} + w_i \lambda_i \end{bmatrix},$$

and

$$D\boldsymbol{\mathcal{F}}(c_i, h_i) = \begin{bmatrix} U_{cc} & U_{ch} + \tau w_i U_{cg} \\ U_{ch} + \tau w_i U_{cg} & U_{hh} + 2\tau w_i U_{hg} + \tau^2 w_i^2 U_{gg} \end{bmatrix}.$$

The Frisch elasticities are the lower components of the vectors in Equations (B33) and (B34), multiplied by w_i/h_i and $(1-\tau)/h_i$ respectively. Specializing to the separable social preferences utility in Equation (4) gives Equations (5) and (6).

B.3 Derivation for program-specific taxes paid

Again, the solution proceeds analogous to Section (B.2). After specializing to the separable utility function

$$U\left(c_i, h_i, g_i, s_i\right) = u\left(c_i, h_i\right) + v\left(g, s_i\right),$$

the elasticities obtained are

$$\epsilon_w^M = \frac{-w_i\left(1-\tau-\sigma\right)\left(\mathfrak{S}_w + \mathfrak{I}\right)}{h_i\left(w_i^2\left(1-\tau-\sigma\right)^2 U_{cc} + 2U_{ch}w_i\left(1-\tau-\sigma\right) + U_{hh} + U_{gg}\tau^2 w_i^2 + 2U_{gs}\tau\sigma w_i^2 + U_{ss}\sigma^2 w_i^2\right)},$$

$$\epsilon_{1-\sigma}^M = \frac{-w_i\left(1-\sigma\right)\left(\mathfrak{S}_{1-\sigma} + \mathfrak{I}\right)}{h_i\left(w_i^2\left(1-\tau-\sigma\right)^2 U_{cc} + 2U_{ch}w_i\left(1-\tau-\sigma\right) + U_{hh} + U_{gg}\tau^2 w_i^2 + 2U_{gs}\tau\sigma w_i^2 + U_{ss}\sigma^2 w_i^2\right)},$$

$$\epsilon_w^H = \frac{-w_i\left(1-\tau-\sigma\right)\left(\mathfrak{S}_w - \left(\frac{\tau}{1-\tau-\sigma}\frac{U_g}{U_c} + \frac{\sigma}{1-\tau-\sigma}\frac{U_s}{U_c}\right)\mathfrak{I}\right)}{h_i\left(w_i^2\left(1-\tau-\sigma\right)^2 U_{cc} + 2U_{ch}w_i\left(1-\tau-\sigma\right) + U_{hh} + U_{gg}\tau^2 w_i^2 + 2U_{gs}\tau\sigma w_i^2 + U_{ss}\sigma^2 w_i^2\right)},$$

$$\epsilon_{1-\sigma}^H = \frac{-w_i\left(1-\sigma\right)\left(\mathfrak{S}_{1-\sigma} + \frac{U_s}{U_c}\mathfrak{I}\right)}{h_i\left(w_i^2\left(1-\tau-\sigma\right)^2 U_{cc} + 2U_{ch}w_i\left(1-\tau-\sigma\right) + U_{hh} + U_{gg}\tau^2 w_i^2 + 2U_{gs}\tau\sigma w_i^2 + U_{ss}\sigma^2 w_i^2\right)},$$

$$\epsilon_w^F = \frac{-w_i\left(1-\tau-\sigma\right)U_{cc}\mathfrak{S}_w}{h_i\left(U_{cc}U_{hh} + U_{cc}U_{gg}\tau^2 w_i^2 + 2U_{cc}U_{gs}\tau\sigma w_i^2 + U_{cc}U_{ss}\sigma^2 w_i^2 - U_{ch}^2\right)},$$

and

$$\epsilon_{1-\sigma}^F = \frac{-w_i\left(1-\sigma\right)U_{cc}\mathfrak{S}_{1-\sigma}}{h_i\left(U_{cc}U_{hh} + U_{cc}U_{gg}\tau^2 w_i^2 + 2U_{cc}U_{gs}\tau\sigma w_i^2 + U_{cc}U_{ss}\sigma^2 w_i^2 - U_{ch}^2\right)},$$

where

$$\mathfrak{S}_w = U_c + \frac{\tau}{1-\tau-\sigma}\left(U_g + U_{gg}g_i\right) + \frac{\sigma}{1-\tau-\sigma}\left(U_s + U_{ss}s_i\right) + 2\frac{\tau\sigma}{1-\tau-\sigma}U_{gs}h_i w_i,$$

$$\mathfrak{S}_{1-\sigma} = U_c - U_s - U_{ss}s_i - U_{gs}g_i,$$

and

$$\mathfrak{I} = U_{cc}h_i w_i\left(1-\tau-\sigma\right) + U_{ch}h_i.$$

B.4 Optimal income taxation

In this subsection, we derive implications of government-regarding social preferences for optimal income taxation. We examine two main cases: the optimal uniform tax rate and the optimal nonlinear tax schedule. Derivations below follow the steps in Piketty and Saez (2013).

B.4.1 Optimal uniform income taxation

Consider a uniform tax rate τ used to fund a demogrant R (a transfer of equal value for all individuals). The government chooses τ to maximize[33]

$$SWF = \int_i \omega_i \Gamma\left(\hat{V}_i\left(\tau, R\right)\right) d\nu\left(i\right), \tag{B35}$$

where the integration is over a measure-one population of individuals in the economy (indexed by i) with distribution $\nu\left(i\right)$, ω_i is a set of Pareto weights, $\Gamma\left(\cdot\right)$ is increasing and concave. $\hat{V}_i\left(\tau, R\right)$ is the i-specific intratemporal value function from Section B.2:

$$\begin{aligned}\hat{V}_i\left(\tau, R\right) &\equiv V\left(a_i, w_i, \tau, N_i + R\right)|_{A_i} \\ &= U\left(A_i + \left(1-\tau\right) w_i h_i^* + N_i + R, h_i^*, \tau w_i h_i^*\right) + \beta E\left[\cdot\right], \end{aligned} \tag{B36}$$

where the $\beta E\left[\cdot\right]$ term is a constant once we hold A_i constant, and $h_i^* = h_i\left(a_i, w_i, \tau, N_i + R\right)|_{A_i}$. The demogrant R is evenly distributed to everybody, so

$$R = \tau \int_i z_i^* d\nu\left(i\right) - E \tag{B37}$$

where E is some government non-transfer spending and $z_i^* \equiv w_i h_i^*$ is total pre-tax income, introduced to save on some notation. Note that h_i^* (and hence z_i^*) is a function of τ, so

$$\begin{aligned}\frac{dR}{d\tau} &= \int_i z_i^* d\nu\left(i\right) + \tau \int_i w_i \frac{\partial h_i^*}{\partial \tau} d\nu\left(i\right) \\ &= \int_i z_i^* d\nu\left(i\right) - \frac{\tau}{1-\tau} \epsilon_{1-\tau}^M \int_i z_i^* d\nu\left(i\right) \end{aligned} \tag{B38}$$

after some derivations. As is well known in the literature, when the tax rate increases, the demogrant increases mechanically (the first term) but not as much as expected because labor supply decreases (the second term), the latter of which is governed by the Marshallian elasticity $\epsilon_{1-\tau}^M$.

The first-order condition for the government, computed by differentiating Equation (B35) subject to Equation (B37), is

$$\int_i \omega_i \Gamma'\left(\hat{V}_i\left(\tau, R\right)\right) \left[U_{c,i}\left[-z_i^* + \frac{dR}{d\tau}\right] + U_g z_i^*\right] d\nu\left(i\right) = 0. \tag{B39}$$

[33] Following the literature, we introduce the i notation to emphasize possible heterogeneity in the population. All other notational elements are chosen to maintain continuity with previous sections.

Note that $U_g z_i^*$ does not envelope out since individuals cannot choose the amount of government expenditure. For comparative statics, it is convenient to use the "usual" social marginal welfare weight on person i (usual because it does not incorporate any U_g term):

$$\gamma_i \equiv \frac{\omega_i \Gamma' \left(\hat{V}_i \left(\tau, R \right) \right) U_c}{\int_j \omega_j \Gamma' \left(\hat{V}_j \left(\tau, R \right) \right) U_{c,j}}. \tag{B40}$$

Substituting Equations (B38) and (B40) into Equation (B39), and after some manipulation, we get

$$\tau = \frac{1 - \bar{\gamma}}{1 - \bar{\gamma} + \epsilon^M_{1-\tau}} \tag{B41}$$

where

$$\bar{\gamma} \equiv \frac{\int_i \gamma_i \left(1 - \frac{U_g}{U_c} \right) z_i^* d\nu \left(i \right)}{\int_i z_i^* d\nu \left(i \right)} \tag{B42}$$

is a penalized income-weighted average of the social marginal welfare weight, with penalty due to government-regarding social preferences.

Compared to a world without social preferences, the only difference is that $\bar{\gamma}$ in Equation (B42) now has $\left(1 - \frac{U_g}{U_c} \right)$ instead of 1 in the numerator. Hence, assuming that $U_g \geq 0$ for all individuals, the existence of government-regarding social preferences lead to smaller $\bar{\gamma}$.[34] From Equation (B41), this means that the optimal tax rate would be larger in the presence of positive social preferences.

[34] $U_g \geq 0$ for all individuals is of course an overly-conservative bound, since the presence of some individuals with negative social preferences need not necessarily turn the whole average negative. The precise bound can be obtained by rewriting Equation (B42) as

$$\bar{\gamma} = E \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]} \right] \left(1 - \left(E \left[\frac{U_g}{U_c} \right] + \frac{Cov \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]}, \frac{U_g}{U_c} \right]}{E \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]} \right]} \right) \right).$$

The first term $E \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]} \right]$ is the non-penalized weighted average social marginal welfare weight (i.e. the $\bar{\gamma}$ in a world without social preferences). Hence, $\bar{\gamma}$ decreases relative to a world without social preferences if $E \left[\frac{U_g}{U_c} \right] + \frac{Cov \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]}, \frac{U_g}{U_c} \right]}{E \left[\gamma_i \frac{z_i^*}{E \left[z_i^* \right]} \right]} > 0$.

B.4.2 Optimal nonlinear tax schedule

Now the government's problem is to choose a nonlinear tax schedule $T(z)$ to maximize

$$SWF = \int_i \omega_i \Gamma\left(\hat{V}_i(T,R)\right) d\nu(i) \tag{B43}$$

$$\text{s.t. } R = \int_i T(z_i^*)\, d\nu(i) - E$$

where

$$\hat{V}_i(T,R) = V(a_i, w_i, \tau_i, N_i + R)\,|_{A_i} = U(A_i + z_i^* - T(z_i^*) + N_i + R, h_i^*, T(z_i^*)) + \beta E[\cdot].$$

$T(z)$ gives the total amount of taxes paid, such that $T'(z)$ is the marginal tax rate at any income z, and $\hat{V}_i(\cdot)$ is a higher order function for notational simplicity. We assume away income effects, and follow the graphical argument around Piketty and Saez (2013, Figure 6) to consider the mechanical, behavioral, and welfare loss effects of a $d\tau$ increase in the marginal tax rate between z and dz.

The first two effects do not depend on the utility function and retain their standard forms (well known in the literature). The mechanical effect is the total increase in tax rate $dzd\tau$ applied to all incomes above z, hence:

$$dM = dzd\tau\,[1 - H(z)], \tag{B44}$$

where $H(z)$ is the (endogenous) cumulative distribution function of income with density $h(z)$. The behavioral effect only affects the $h(z)\,dz$ individuals with incomes between z and dz since all other incomes face no change in the marginal tax rate and we have assumed away income effects. An affected individual i changes her income by $w_i \frac{\partial h_i^*}{\partial \tau} d\tau = -z_i^* \frac{d\tau}{1-T'(z)} \epsilon_{1-\tau,i}$, where $\epsilon_{1-\tau,i}$ is the individual-specific ETI, yielding tax revenue consequences of $-z_i^* \frac{d\tau}{1-T'(z)} \epsilon_{1-\tau,i} T'(z)$. Put together, the behavioral effect is

$$dB = -dzd\tau h(z)\,\epsilon_{1-\tau} z \frac{T'(z)}{1-T'(z)}. \tag{B45}$$

The welfare loss of the change is obtained by considering the effect on the social welfare function in Equation (B43), restricted to the people with incomes above z who pay $dzd\tau$ more in taxes, and noting that any labor supply effects envelope out when considering the effects on welfare. Using the marginal value of public funds for the government p to convert

from utility to dollars, the money-metric welfare loss is (after some algebraic manipulation)

$$dW = -dzd\tau \int_{z_i>z} \gamma_i \left(1 - \frac{U_g}{U_c}\right) dH\left(z_i\right) \tag{B46}$$

$$= -dzd\tau \left[1 - H\left(z\right)\right] \gamma^+\left(z\right), \tag{B47}$$

where $\gamma_i \equiv \omega_i \Gamma' \left(\hat{V}_i\left(T, R\right)\right) U_c/p$ in Equation (B46) is again the "usual" social marginal welfare weight on person i, and in Equation (B47),

$$\gamma^+\left(z\right) \equiv \frac{1}{1 - H\left(z\right)} \int_{z_i>z} \gamma_i \left(1 - \frac{U_g}{U_c}\right) dH\left(z_i\right) \tag{B48}$$

is the average social marginal welfare weight for individuals with income above z, penalized by government-regarding social preferences.

The optimal marginal tax rate at any z is obtained by setting the sum of the effects in Equations (B44), (B45), and (B47) to zero. This yields

$$T'\left(z\right) = \frac{1 - \gamma^+\left(z\right)}{1 - \gamma^+\left(z\right) + \alpha\left(z\right)\epsilon_{1-\tau}}$$

where $\alpha\left(z\right) \equiv \frac{zh(z)}{[1-H(z)]}$ is a parameter that is related to the Pareto distribution.

Again, compared to a world without social preferences, the only difference is that $\gamma^+\left(z\right)$ in Equation (B48) incorporates a social preferences penalization term. On top of the interpretation from Section B.4.1—positive social preferences allows the government to levy higher income taxes on average—Equation (B48) reveals that heterogeneity in social preferences can also affect the marginal tax rate at different points of the income distribution.

C Survey and vignette experiment details

C.1 Ensuring high-quality response

We maximized response quality in several ways.

In our survey landing page (consent page), we followed a good practice in the literature and warned participants that responding without adequate effort might result in their responses being marked as low quality and not used (see e.g. Stantcheva 2023). Also following the literature, we appealed to their sense of social responsibility by emphasizing that their responses might form the basis for policy recommendations, and requested that they read the questions carefully and answered honestly. To avoid selection based on survey topics, our

landing page provided limited information about what the survey was about. (Specifically, "We are a group of independent non-partisan researchers studying how people around the world make decisions about their jobs.") Throughout the survey, we used simple and neutral language to avoid giving respondents the impression that we had a hypothesis in mind.

We took several steps to ensure accurate measurement of employment variables. For ease of answering, when asking respondents about their earnings, we adapted our questions to the respondents by offering the option of entering an annual, monthly or weekly figure. Because the hourly wage (earnings divided by work hours) can be difficult to measure (see e.g. Borjas 1980), we programmed a dynamic check of the wage in the survey instrument. Specifically, immediately after the respondent entered her earnings and work hours, we showed her the computed wage and asked her if it sounded "about right". If the answer was "No", we provided the computation formula and requested that she checked the figures again. The key employment variables (work hours, earnings, and verified wage) were all asked on the same webpage to make internal consistency easy for the respondent (since there was no need to click the "back" button to change numbers).

Our vignette experiment was necessarily more complex than other parts of our survey which comprised simple-to-answer questions. We implemented four features to ensure the highest possible quality answers. First, before the vignette experiment, we required respondents to go through an animated slideshow that walked them through the instructions. In total, there were ten slides that explained the hypothetical scenarios, the assumptions (as described in Section 3.2), and the choices they would be asked to make. Sentences on the slides appeared at 250 words per minute (between the usual reading and speaking speed), and the next slide button was grayed out (i.e., not responsive) until all the text had appeared. Hence, unlike an instructional video, participants could not simply start the slideshow and return after some time to resume the survey—they had to advance the slides a total of ten times throughout the slideshow.

Second, we focused the respondents' attention on the question at hand by hiding all non-relevant information in the questionnaire matrix. For example, when the respondent had to choose whether to work or not in the first scenario, we hid the second scenario and the question on which scenario he preferred. When he had to make the analogous choice for the second scenario, we similarly hid the first scenario, including the previous response to minimize anchoring effects. We revealed all scenarios and choices only when the respondent was to choose between the two scenarios.

Third, on the first pair of scenarios, we ensured that respondents understood the layout and how to answer by showing an interactive "guided website tour" through each component (see Appendix Figure A1). Every time we introduced a change in the vignette layout after-

wards (i.e., pairs 5 and 9), we showed a new interactive guided tour to explain the change and to maintain their attention.

Fourth, we programmed a check for answers exhibiting intransitive preferences—choices to work in a scenario of a pair and not the other, but yet the respondent preferred the latter scenario to the former—and showed a non-intrusive prompt asking respondents to double-check their answers when an intransitive choice was made. Without the prompt, the worry is that some (more attentive) respondents may notice that we allow inconsistent choices, think that the survey was badly designed, and lose motivation. We kept track of when the prompt was observed, and investigate robustness to observation of the prompt in Section 4.2.

When eliciting opinions about government spending on specific programs, we used a more complex sampling mechanism (of the programs) to balance multiple objectives. Our primary goal is to study how opinions of a program affects the program-specific wage-tax elasticity wedge in the pairs 9 and 10 of vignettes (one program per pair). These two programs were selected from a list of ten using a randomization procedure that weighted dissimilar programs more. Specifically, each pair of programs (e.g. Social Security and Medicare) had a probability of selection that was proportional to the Euclidean distance in liking score between the programs of the pair.[35] Later on in the survey when we asked respondents how much they liked/disliked their tax money being used to fund specific programs, each respondent was asked to rate five programs including the two they saw in pairs 9 and 10 of the vignette and three other randomly chosen programs. We added the three randomly chosen programs to better mask our objective for asking about the opinion of government spending and because these opinions are of general interest in and of itself. (We did not present the full list of ten programs to every respondent to reduce response burden.)

We conducted the survey when there was some worry about bots being used to answer surveys (Goodrich et al. 2023). Thus, we followed good practices in the literature to ensure that our respondents are legitimate and verified (Stantcheva 2023). Besides distributing our survey through a reliable company which also performed checks of their participants, we incorporated Qualtric's bot detection capabilities and added a "honey pot" question on our landing page. The question "Check this box if it applies to you. [I am a piece of software/I am not a piece of software]" appeared for one second and was displaced downwards by one screen height, so that a human participant saw a blank screen for the first second (consistent with the webpage still loading). Among all participants who started the survey, this question only detected one bot that passed upstream bot checks. (We screened the respondent out.)

We kept track of three measures of response quality in the survey. First, we followed

[35]The Euclidean distances were computed based on a pilot dataset. Interpretation of these distances as a measure of dissimilarity is frequently used in data clustering algorithms.

the literature (see e.g. Mas and Pallais 2017) to build two attention checks into the survey. The first check (which we refer to as the pre-vignette attention check) doubles as a comprehension check and asked participants a simple question about the instruction of the vignette following the slideshow. Specifically, respondents were asked whether the income tax rate in the hypothetical scenarios referred to federal, state or local income tax rate (the correct answer being federal). The second check (the in-vignette attention check) was built into pair 7 of our our vignette experiment. In this pair, by design, one of the two scenarios was strictly dominating: both scenarios showed the same tax rate and work hours but one had a strictly higher hourly wage rate. We would expect attentive participants to prefer the strictly dominating scenario. We chose pair 7 as our in-vignette attention check since vignette fatigue would likely to affect respondents the most around this pair.

Second, we exploited the interactive nature of our instructional slideshow to identify "impatient" respondents. Since respondents could not advance the slides until all the text were displayed—and the graying out of the "next" button made this obvious—respondents who were thinking about the instructions carefully would only click "next" a total of ten times. Hence, the number of times this "next" button was clicked when it was still grayed out serves as a natural measure of response quality. In practice, we labeled respondents who clicked more than twenty times (twice as many as needed) as impatient respondents.

Finally, the the time spent by respondents on the survey as a whole, as well as on the vignette experiment and other questions, allow us to identify respondents who spent too little time answering questions. This is a standard measure used in the survey research and practitioner literature (Zhang and Conrad 2014; Greszki et al. 2015; Leiner 2019). The literature and practitioners often use thresholds relative to the median time (see e.g. Greszki et al. 2015); we follow this practice and label respondents who responded faster than half the median time as "too fast". We considered three time intervals when setting this measure: the overall time taken, the time taken on the vignette experiment (our main focus), and the time taken for the government expenditure liking questions (to measure quality near the end of the survey, and because we use a quality sample when investigating the impact of liking the program that taxes fund).

C.2 Sample

Our target sample comprised US residents aged 25–64 who were working at the time of the survey. Full-time students who might have worked part-time were excluded from the survey since they might be less familiar with the labor market implications of their choices. We also excluded the self-employed (including unpaid family workers) as they might find it more

Appendix Table C1: Sample size and data-cleaning steps

	Number of Respondents
Started the survey and not screened out	12,164
Got to just before the instructional slideshow	8,338
Passed the instructional slideshow	7,856
Completed the vignette experiment	7,158
Complete the entire survey	7,122
Excluded:	
Work hours, consumption, or tax amounts below 5th or above 95th percentiles	1,611
Vignette consumption not imputable	90
Skipped slide show	45
Final Sample	5,440

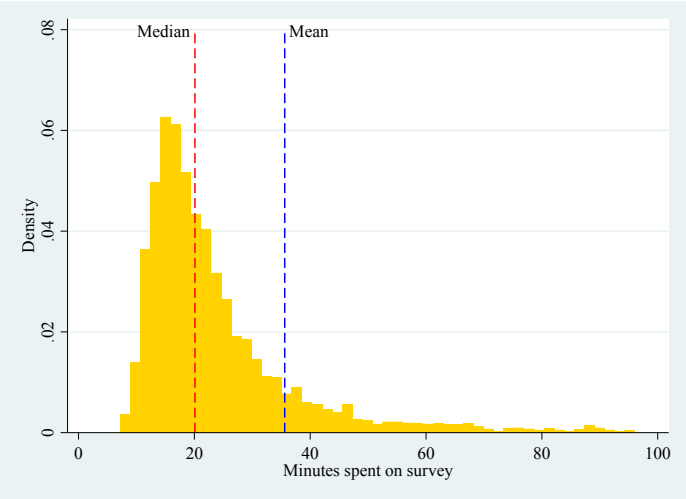
Notes: This table shows the number of respondents who started our survey and the sample size reduction for each of our data-cleaning steps. Respondents may be excluded for more than one of the three reasons stated.

difficult to relate to the scenarios in our vignette which were about finding a new job with similar conditions to the participant's current job.We asked Bilendi/Respondi, our survey company, to target participants such that our sample is representative along the age, gender, and income group dimensions. (We provided target proportions for different groups based on the working population in the 2022 CPS data.)

Our survey started with demographic questions to determine participants' eligibility: those who did not meet our sample definition were screened out of the survey. 12,164 respondents started the survey and were not screened out. Appendix Table C1 describes, step-by-step, the sample size reductions for each of our data cleaning steps to arrive at our main sample. Appendix Figure C1 shows the distribution of the time respondents in our main sample spent on completing the survey. The mean and median duration was 35.6 and 20.1 minutes, respectively.

We flag respondents with potentially low-quality answers and exclude them in some of our analyses. We did this using the three measures—"inattentive", "impatient" and "too fast"—described in Appendix C.1. Appendix Table C2 reports the number of these respondents and their proportion of our main sample.

Appendix Figure C1: Time Spent on Survey



Notes: The figure shows the distribution of the time respondents in the main analysis sample spent on the survey (truncated at 100 minutes). The mean duration is 35.6 minutes, the median 20.1, and the 25th and 75th percentile are 15.5 and 28.6, respectively.

Appendix Table C2: Inattentive, impatient or too fast respondents

	Number of respondents	Percent of final sample
Inattentive in pre-vignette attention check	381	7.0
Inattentive in in-vignette attention check	447	8.2
Inattentive in either attention check	726	13.3
Clicked next slide >20 times during vignette instruction slideshow	909	16.7
Time taken for any of overall, vignette, and gov. expenditure liking questions is less than half the median time	1,286	23.6

Notes: This table shows the number of respondents and the proportion of the main sample that are "inattentive" (failed our pre-vignette or in-vignette attention checks), "impatient" (clicked next more than 20 times during the vignette instructional slideshow, and "too fast" (those with time spent on any of the overall, vignette and government expenditure liking questions less than half of the median time.

C.3 Income tax rate simulation

To avoid showing tax rates that differed too much from the respondent's actual situation, our vignette experiment randomly selected tax rates around the respondent's actual tax rate. The actual tax rate was computed by simulation, both to simplify the details that respondents should focus on (e.g. we simply needed them to focus on the average tax rate for our exercise, ignoring the complications that come with marginal tax rates) and because respondents might not recall the tax rates that they were paying. To avoid the internet speed reduction and risk of connection failure that comes with querying an external website, we elected to program the simulation directly in the survey, instead of obtaining the tax rate from an external source like the National Bureau of Economic Research TAXSIM.

We modeled three filing categories: married filing jointly if the respondent was married, single-filer if the respondent was non-married without any children, and head of household if the respondent was non-married with children. Total household income comprised wage income, interest and dividend income, capital gains income, and business income.[36] Wage income was directly collected in our survey (for both the respondent and their spouse if present). We imputed the other types of income separately for married and single respondents. To do this, we ran a regression of each income type on several demographic characteristics using the 2015–2019 Current Population Survey data. We then used these models and the demographic information collected in our survey to predict the interest and dividend income (set to zero if the resulting prediction was negative), capital gains income and business income for each respondent.

We modeled deductions as a weighted average of the standard deduction and an itemized deduction which varied across household income and filing category. The weights and itemized deduction amount were estimated based on the Internal Revenue Service's Statistics of Income (Internal Revenue Service 2021, Table 1.2). We computed taxable income as total household income less deductions, and applied the 2022 federal income tax brackets to obtain an income tax before credits amount for each respondent.

Next, we modeled two large tax credits, the Child Tax Credit and the Earned Income Tax Credit, using the federal tax credit formulas which consider income, number of children, and marital status. We chose these two credits to model because they were large credits that were likely to apply to our respondent sample and did not require additional questions that deviated from our main research question. We then computed the simulated income tax for each respondent as the income tax before credits less the imputed tax credits. Finally, we divided the income tax after credits by total household income to obtain the average tax

[36] All numbers in the income tax rate simulation were adjusted to be in 2022 dollars.

rate for each respondent.

D Full questionnaire

A live version of the survey is available at https://nus.syd1.qualtrics.com/jfe/form/SV_cAszYDfScYSrEUu. The full questionnaire is available at https://www.pinchuanong.com/wp-content/uploads/2024/07/appx-d.pdf. Some of our background and political view questions were adapted from Stantcheva (2022) and Doherty et al. (2015).