NBER WORKING PAPER SERIES

DE GUSTIBUS NON EST TAXANDUM: THEORY AND EVIDENCE ON PREFERENCE HETEROGENEITY AND REDISTRIBUTION

Benjamin B. Lockwood
Matthew C. Weinzierl

Working Paper 17784
http://www.nber.org/papers/w17784

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2012

We are grateful to Robert Barro, Rafael di Tella, Alex Gelber, Caroline Hoxby, Louis Kaplow, Narayana Kocherlakota, Erzo F.P. Luttmer, Greg Mankiw, David Moss, Eric Nelson, Julio Rotemberg, Danny Shoag, Aleh Tsyvinski, Glen Weyl, Danny Yagan, and seminar participants at Harvard and Michigan for helpful comments and suggestions on earlier versions of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

De Gustibus non est Taxandum: Theory and Evidence on Preference Heterogeneity and Redistribution
Benjamin B. Lockwood and Matthew C. Weinzierl
NBER Working Paper No. 17784
January 2012
JEL No. H2,H21

ABSTRACT

Preferences over consumption and leisure play no role in the standard optimal tax model, which attributes all variation in earnings to differences in income-earning ability. We show how to incorporate these preferences, which like ability are publicly unobservable, into the standard model in a tractable way. In this more general model, the policy designer must guess at the relative importance of ability and preferences in explaining variation in earnings. We show that such preferences could, in principle, increase or decrease optimal redistribution. In the most plausible specifications of the model, however, the result is clear: greater variation in preferences lowers the optimal extent of redistribution. To generate more redistribution than in standard results, one must assume that the desire for income is inversely related to income earned. This result holds even when the conventional model accurately describes the average individual, and it suggests one potential resolution to the puzzle of why observed redistribution is in some cases weaker than conventional theory would suggest. We then establish a new empirical finding that confirms this model's central policy prediction across developed countries and U.S. states. In countries and states with more heterogeneous tastes for consumption relative to leisure, redistribution is statistically significantly lower.

Benjamin B. Lockwood
Littauer Center, Room 200
Harvard University
Cambridge, MA 02138
lockwood@fas.harvard.edu

Matthew C. Weinzierl
Harvard Business School
277 Morgan
Soldiers Field
Boston, MA 02163
and NBER
mweinzierl@hbs.edu

Introduction

Individuals differ in the value they place on consumption relative to leisure. These preference differences help explain why some earn more than others, and they are a central part of popular and scholarly debates over taxation. In this paper, we show that variation in these preferences may also help explain why the extent of redistribution varies across countries and U.S. states, and why (at least in the case of the United States) redistribution is weaker than conventional theory would suggest. More generally, we argue that neglecting the role of preferences substantially impairs our understanding of both optimal and existing tax policy.

Surprisingly, such preference differences are absent from conventional optimal tax theory. That theory, instead, attributes all variation in earnings to differences in income-earning ability. We generalize the conventional model by allowing the tax designer to attribute observed variation in incomes to ability, preferences, or a mixture of the two. The tax designer maximizes social welfare by choosing a universal grant and a linear tax rate, a simplified tax system that provides us with a transparent measure of redistribution. We use analysis of an illustrative version of this model and extensive numerical simulations of the full model to characterize the effects of preferences on redistribution.[1]

We derive novel, clear results from our generalized model. We show that variation in preferences can, in principle, increase or decrease the optimal extent of redistribution. In the most plausible specifications of the model, however, greater variation in preferences lowers optimal redistribution. To generate more redistribution than in the conventional model, one would have to assume (counterintuitively) that the desire for income is inversely related to income earned: i.e, that high earners have a lower marginal rate of substitution of consumption for leisure than low earners at the same starting levels of consumption and leisure. We show these results first for a model in which preferences and ability are the same for all individuals with a given income (as in the standard model). In that setting, attributing more of earnings variation to preferences rather than ability lowers the optimal linear tax rate. Then, we show that if preferences and ability vary conditional on income, preference heterogeneity is even more likely to reduce optimal redistribution relative to standard results. In particular, suppose that the average individual with a given income has the same profile of preferences and ability as assumed in the conventional model. If other individuals with the same income have different profiles, then optimal redistribution falls.

The intuition for these results is that income is a worse signal of ability when preferences play a greater role in driving earnings variation. Redistribution aimed at helping those with low ability will be less accurate: some high earners will have low ability and some low earners will have high ability. A tax designer weighing the distortionary costs of redistribution against its smaller benefits will choose a lower optimal level.

These results suggest that the conventional Mirrleesian assumption is a special one that is, at least in this way, likely to overstate the optimal extent of redistribution. We thereby provide one possible explanation for the finding in Diamond and Saez (2011) that, based on the standard model, "Very high earnings should be subject to rising marginal rates and higher rates than current U.S. policy for top earners."

Finally, we present a new empirical finding that suggests our model has descriptive, not just prescriptive power. Using measures of redistribution and individuals' responses to survey questions on how much they value leisure and material goods, we confirm a direct prediction of our model: more variance in reported preferences is significantly associated with less redistribution across both countries and U.S. states, conditional on observed variation in incomes and the correlation of income with preferences. These results are robust

[1] We are focused on a different type of preference variation than that which bears on the optimality of uniform commodity taxation, as in, e.g., Saez (2002), Diamond and Spinnewijn (2011), and Golosov, Troshkin, Tsyvinski, and Weinzierl (2011).

to controlling for a number of observable variables, including average income and ethnic fractionalization, and they suggest a new aspect of the interpretation of the much-studied differences in redistribution across countries.

The paper proceeds as follows. Section 1 briefly summarizes how this paper relates to the conventional model's treatment of preferences and to prior work on relaxing that approach. Section 2 describes a tractable and flexible way to include preferences in a generalized version of that conventional model. Section 3 stays close to the standard model, limiting heterogeneity to one dimension but allowing the tax designer to interpret that heterogeneity as a combination of preferences and ability. Section 4 considers the case with conditional variation in preferences and ability. Section 5 contains our empirical analysis of the relationship between preference heterogeneity and redistribution. Section 6 concludes.

1 This analysis in the context of prior related work

Heterogeneity in preferences was assumed away in the conventional optimal tax model because of concerns over how to translate ordinal preferences into the cardinal utility functions necessary for using a Utilitarian social objective function.[2] As Mirrlees (1971) acknowledged, his formulation of the optimal tax problem makes some strong assumptions, the second of which is:

> "Differences in tastes, in family size and composition, and in voluntary transfers, are ignored. These raise rather different kinds of problems, and it is natural to assume them away."

This simplification freed Mirrlees[3] to assume that the only way in which people differ is in their ability to earn income. His approach has been dominant ever since.

The assumption that all heterogeneity takes the form of ability is a bold one, because allowing heterogeneity to enter in a more general manner can render the Mirrleesian optimal tax model powerless. As Sandmo (1993) showed in an insightful paper, the problem is that observationally equivalent representations of heterogeneity can yield dramatically different optimal tax policies. An individual with low income may have low ability, following Mirrlees' model, or he may place a low value on consumption. In the former case, optimal policy redistributes to him; in the latter, it does not and, under some cardinalizations of preferences, it may even redistribute from him. As long as both ability and preferences are unobservable, it is not possible to distinguish these interpretations with data on economic behavior.

To pin down an optimal policy, we must assume that one of the interpretations, and thus cardinalizations, of heterogeneity is the appropriate one. Building on the work of Sandmo, recent contributions by Boadway et al. (2002) and Kaplow (2008) incorporate heterogeneous preferences while remaining agnostic about the appropriate cardinalization and show that the implications for optimal redistribution are ambiguous. Choné and Laroque (2010) provide an elegant discussion of the same lesson for the sign of marginal income tax rates, which are always nonnegative in the conventional model but may be negative once heterogeneous preferences (or opportunity costs of work, in their words) are incorporated.

[2] The use of a Utilitarian social welfare function in this context has its formal roots in Harsanyi (1953, 1955).

[3] Mirrlees was not the first to adopt this simplification. Pigou (1928) wrote, in a classic text: "Of course, in so far as tastes and temperaments differ, allowance ought, in strictness, to be made for this fact...But, since it is impossible in practice to take account of variations between different people's capacity for enjoyment, this consideration must be ignored, and the assumption made, for want of a better, that temperamentally all taxpayers are alike."

In this paper, we focus our attention on the cardinalization of utility under which preference heterogeneity justifies no redistribution. We show that incorporating such preference heterogeneity does not necessarily imply less optimal redistribution, and we clarify the (plausible) conditions under which it does.

The potential impact of this interpretation of preferences on optimal taxation has long been emphasized by those skeptical of redistribution, consistent with our results. Robert Nozick, an influential modern philosopher and leading expositor of the Libertarian normative framework, wrote in his book *Anarchy, State, and Utopia (1974):* "Why should we treat the man whose happiness requires certain material goods or services differently from the man whose preferences and desires make such goods unnecessary for his happiness?" Along similar lines, Milton Friedman (1962) wrote: "Given individuals whom we are prepared to regard as alike in ability and initial resources, if some have a greater taste for leisure and others for marketable goods, inequality of return through the market is necessary to achieve equality of total return or equality of treatment." The broad influence[4] of arguments such as Nozick's and Friedman's suggests, at least to us, that a convincing theory of optimal taxation ought to carefully address the substance of their critique.

Our approach is thus close in spirit to the important recent work of Fleurbaey and Maniquet (2006), who explore the effects of preference heterogeneity on optimal taxation from the perspective of the social choice literature. As do we, they assume (in the form of so-called fairness assumptions) that the appropriate cardinalization of preferences over consumption and leisure is such that optimal policy redistributes across ability but not across preferences. Unlike us, however, they impose informational constraints on the social planner which rule out conventional Utilitarian social welfare functions and which, in combination with particular fairness requirements on allocations, imply the use of a maximin social welfare function.[5] Our analysis can be seen as a complement to theirs. While their paper argues for and adopts a specific normative approach that sets it apart from the conventional Mirrleesian literature, we build directly on that dominant theory to show the effects of preference heterogeneity on conventional results, and we provide evidence that these effects are consistent with existing policy variation.

A second valuable and recent analysis close to ours is by Judd and Su (2006), who add multiple dimensions of heterogeneity to a standard Mirrleesian model and simulate optimal policy.[6] Though they do not simulate[7] policy with the heterogeneity in preferences for consumption relative to leisure that are the focus of our paper, they consider several other types of heterogeneity, including in labor supply elasticities and basic needs. Their results are broadly consistent with ours, in that they show through illustrative numerical simulations that adding other sources of heterogeneity to a given distribution of income-earning abilities typically reduces redistribution.[8] Several features differentiate our paper from theirs, including our calibration of the model to data, our use of continuous type distributions in the numerical simulations, and our analysis of existing policies as a test of the model's predictions. But there is also a more subtle technical difference. Rather than layering on sources of heterogeneity, our paper starts with observed variation in earnings and analyzes the effects of attributing more or less of that variation to preferences. This approach yields two advantages. First, it clarifies the conditions under which adding preference heterogeneity could increase optimal redistribution

[4] For one recent example of this influence, see Bryan Caplan (2011), who argues for less redistributive taxes based on evidence in Kahneman (2011) that such preference differences are widespread at early ages and correlate with outcomes later in life.

[5] Given that setting, they conclude that the optimal income tax should maximize the subsidies to the working poor: that is, it should be quite redistributive to those with low ability but who exert labor effort.

[6] Judd and Su (2006) also address the computational complications that arise from multiple dimensions of heterogeneity. See also Tarkiainen and Tuomala (2004). We avoid these concerns by restricting tax instruments to the grant and linear tax rate.

[7] The possibility of such variation is included in their theoretical setup, however.

[8] Judd and Su (2006) also find that optimal marginal tax rates need not be bounded below by zero. See also Chone and Laroque (2010).

(i.e., if income and preferences are negatively correlated). Second, it makes the model with preference heterogeneity more directly comparable to the conventional model because both must explain an empirical distribution of earnings, whereas adding preference heterogeneity on top of ability heterogeneity changes the distribution of income and, therefore, complicates the comparison of policies.

Two other recent papers have produced results consistent with those of this paper, though they rely on somewhat different channels. Kocherlakota and Phelan (2009) focus on the implications of policymaker uncertainty about the relationship between individuals' preferences and another, welfare-relevant, dimension of heterogeneity such as wealth. They argue that such uncertainty causes a planner using a maximin objective to avoid redistributive policy that is optimal when no such uncertainty is present. Beaudry, Blackorby, and Szalay (2009) indirectly address preference differences by including in their optimal tax analysis differences in productivity of market and non-market labor effort. They show that the optimal redistributive policy makes transfers to the poor conditional on work.

2 A model of optimal tax with heterogeneous preferences

In this section, we generalize Mirrlees' approach by allowing the tax authority to assign shares of the variation in earnings to both unobserved ability and unobserved preferences over consumption and leisure. Mirrlees' assumption is a particular case of this general model. A key technical advantage of our approach is that we include preferences as a component of observable earnings heterogeneity rather than as an addition to unobservable ability. The latter approach imposes too few restrictions on the relationship between ability and preferences for clear comparisons to conventional results to be derived.

2.1 Individuals

Individuals exert labor effort to earn income. They use that income plus any transfers from the government to pay taxes and purchase consumption. Individuals are heterogeneous, such that for any given tax system, there will be a distribution of pre-tax incomes. The sources of heterogeneity are unobservable.

In the *conventional* setup, individuals differ in their abilities to earn income by exerting labor effort. Individuals are indexed with $i \in \{1, 2, ..., I\}$, where the income-earning ability of individual i is denoted w_i. Individual preferences over consumption c and labor effort $l = y/w$, where y is observable labor income, are represented by a common utility function. One tractable example of such a utility function is:

$$U(c_i, l_i) = \frac{(c_i)^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma}\left(\frac{y_i}{w_i}\right)^{\sigma}, \tag{1}$$

where $\gamma > 0$ determines the concavity of utility from consumption and $\sigma > 1$ affects the elasticity of labor supply.

A convenient feature of this conventional formalization is that data on the distribution of earnings y_i is sufficient to extract the distribution of types w_i, given the tax function that translates earnings into consumption c_i. For example, if there are no taxes, individual i chooses:

$$y_i = w_i^{\frac{\sigma}{\sigma+\gamma-1}}. \tag{2}$$

Once a distribution of w_i is known, optimal policy can be determined. Note that heterogeneity in w_i does not affect the utility functions of individuals, but rather their budget constraints.

We introduce to this conventional model preferences over consumption and labor effort (equivalently, leisure). Type is now defined by a duple $\{w_i, \theta_i\}$, where the taste parameter θ_i multiplies the marginal rate of substitution of consumption for labor effort for individual i. It enters the utility function as follows.

$$U_i(c_i, l_i) = \theta_i \frac{(c_i)^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma}\left(\frac{y_i}{w_i}\right)^{\sigma}. \tag{3}$$

Note that, unlike the conventional utility function in (1), the utility function in (3) is type-specific, because utility as a function of consumption and labor effort now depends on the individual's taste parameter θ_i.

Having these two unobservable dimensions of heterogeneity, w_i and θ_i, presents a problem of observational equivalence. The level of earnings chosen by individual i will be a function of both w_i and θ_i, and no observer can determine the relative importance of each factor from data on economic behavior.[9] For example, if there are no taxes, individual i chooses:

$$y = \left(\theta_i w_i^{\sigma}\right)^{\frac{1}{\sigma+\gamma-1}}. \tag{4}$$

In contrast to expression (2), expression (4) indicates that agents may earn the same income and therefore be observationally equivalent despite having different skills and tastes. As we show below, these two dimensions of heterogeneity have dramatically different implications for redistribution, so determining their roles is essential for optimal policy.

This problem of observational equivalence motivates a technical contribution of our paper: transforming the unobservable variables θ_i and w_i into two new variables, one of which is observed, the other of which is not. We use income (which is observable) as one dimension of heterogeneity, and allow for a second, unobserved dimension that adjusts the relative importance of ability and preferences. This transformation allows us to derive novel results that can be compared directly with conventional results.

Formally, in our generalized model, individuals can be fully described by a duple type denoted $\{\lambda_i, \phi_i\}$. The variable λ_i is defined as equal to the earnings chosen by individual i in the absence of taxes, that is:

$$\lambda_i = \left(\theta_i w_i^{\sigma}\right)^{\frac{1}{\sigma+\gamma-1}}. \tag{5}$$

Taking logs of expression (5), we obtain:

$$\ln \lambda_i = \underbrace{\frac{1}{\sigma+\gamma-1}\ln\theta_i}_{\phi_i \ln\lambda_i} + \underbrace{\frac{\sigma}{\sigma+\gamma-1}\ln w_i}_{(1-\phi_i)\ln\lambda_i}. \tag{6}$$

As indicated in (6), we use the variable ϕ_i to divide variation in earnings into two components equal to the two terms on the right-hand side of (6).

We will refer to ϕ_i, somewhat loosely, as the share of income variation attributed to preferences rather than ability, as we can write:

$$\theta_i = \lambda_i^{(\sigma+\gamma-1)\phi_i}, \tag{7}$$

$$w_i = \lambda_i^{(\sigma+\gamma-1)(1-\phi_i)/\sigma}. \tag{8}$$

[9] A corollary to this issue, not pursued in this paper, is that simulations of optimal policy that extract an ability distribution from the income distribution under the assumption that preferences are homogeneous are potentially misleading. See Figure 4 below for a related point.

Thus, the utility function (3) for individual i can be rewritten as:

$$U_i(c_i, l_i) = \lambda_i^{(\sigma+\gamma-1)\phi_i} \frac{(c_i)^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma} \left(\frac{y_i}{\lambda_i^{(\sigma+\gamma-1)(1-\phi_i)/\sigma}} \right)^{\sigma}, \tag{9}$$

This is the utility representation we will use throughout the paper. Note that the only unobservable variable in (9) is ϕ_i, the share of income variation attributed to preferences rather than ability.

The scalars ϕ_i in expression (9) are the key parameters of this model, because we can vary them to consider a wide range of assumptions on the sources of individual heterogeneity. For instance, the conventional Mirrleesian approach boils down to the following assumption:

$$\phi_i = \phi = 0 \text{ for all } i \in \{1, 2, ..., I\}, \tag{10}$$

that is, that the share of income variation attributed to preferences is zero. By assuming (10), Mirrlees' approach sets $\theta_i = 1$ for all $i \in \{1, 2, ..., I\}$, so that preferences over consumption relative to labor effort are the same for all individuals. This paper is about the possibility that $\phi_i \neq 0$ for all or some $i \in \{1, 2, ..., I\}$.

2.2 The tax design problem

A tax designer takes individual behavior as given and sets policy to maximize a social welfare function subject to economic feasibility. We restrict policy to a linear income tax rate τ and a uniform grant g, a rough but commonly-used approximation of optimal tax results in general (see Mirrlees 1971 and Mankiw, Weinzierl, and Yagan 2010). There is a long tradition of considering optimal linear tax functions for the sake of tractability and clarity (e.g., Sheshinksi 1972 and Hellwig 1986), a tradition that has carried through to modern analyses as well (e.g., Farhi and Werning 2011). The most important advantage of doing so for our purposes is that the optimal linear tax rate in this policy can be used as a succinct measure of redistribution, the focus of our analysis.[10]

The social welfare function is weighted Utilitarian. The tax designer applies type-specific multiplicative weights, denoted α_i, to utilities, where $\alpha_i \geq 0$ for all $i \in \{1, 2, ..., I\}$. These weights are used to normalize preferences to be neutral with respect to redistribution.[11] We define this normalization formally as part of the policy problem.

The policy problem can be stated as follows.

Tax Design Problem

$$\max_{\tau, g} \sum_{i=1}^{I} \alpha_i \left[\lambda_i^{(\sigma+\gamma-1)\phi_i} \frac{(c_i)^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma} \left(\frac{y_i}{\lambda_i^{(\sigma+\gamma-1)(1-\phi_i)/\sigma}} \right)^{\sigma} \right], \tag{11}$$

where

$$c_i = g + (1-\tau)\, y_i, \tag{12}$$

subject to feasibility:

$$\sum_{i=1}^{I} (\tau y_i - g) \geq 0, \tag{13}$$

[10] A natural extension of this paper is to consider the effect of preferences on optimal nonlinear taxes.

[11] See Section 1's discussion of the related fairness requirements of Fleurbaey and Maniquet (2006) and the claims by Nozick and Friedman.

where each individual chooses income optimally given the tax system, so that the first-order condition

$$(1-\tau)\,\lambda_i^{(\sigma+\gamma-1)}\,(c_i)^{-\gamma} - (y_i)^{\sigma-1} = 0 \tag{14}$$

holds for all $i \in \{1, 2, ..., I\}$*, and where the welfare weights* $\{\alpha_i\}_{i=1}^{I}$ *are chosen so that the Tax Design Problem satisfies* Preference Neutrality, *defined as follows:*

Definition: Preference Neutrality

Preference Neutrality is satisfied by the Tax Design Problem if and only if the following condition holds:

$$\phi_i = 1 \textit{ for all } i \in \{1, 2, ..., I\} \Rightarrow \tau = 0 \textit{ and } g = 0. \tag{15}$$

That is, if all individuals have the same ability (i.e., $w_i = \bar{w}$ *for some constant* $\bar{w}$*), then the solution to the Tax Design Problem entails no redistribution of income.*

The assumption of Preference Neutrality requires that in a world with differences only in preferences, no interpersonal transfers are justified. By assuming Preference Neutrality, we pin down values for the welfare weights that can then be applied to the more general situation with heterogeneity in both preferences and ability.[12]

Note that Preference Neutrality formalizes an important conceptual distinction between the two unobservable dimensions of heterogeneity in this paper, w_i and θ_i. While w_i is entirely deserving of redistribution, θ_i is entirely undeserving of redistribution. This binary distinction could be generalized to include intermediate cases. That is, we might array the potentially large number of ways in which individuals differ along a spectrum, the poles of which are w_i and θ_i. For instance, the presence of dependent children can be thought of in part as generating greater consumption needs (which may justify redistribution) and in part as being a choice variable of the parents (which would not justify redistribution). Importantly, even such intermediate cases would have the same qualititative implications for redistribution as does θ_i, as long as they cannot be taxed directly, because the conventional model assumes all heterogeneity is best modeled as of the w_i form.

In the next two sections we study the tax design problem under two assumptions about the nature of preference variation: i.e., whether preferences are uniform conditional on income or whether they vary conditional on income.

3 Optimal policy with $\phi_i = \phi$ for all types

We start our analysis with a specific case of the Tax Design Problem that keeps us close to the conventional Mirrleesian approach but that nevertheless generates dramatically different results. In particular, in this section we assume $\phi_i = \phi$ for all $i \in \{1, 2, ..., I\}$, meaning that all individuals with a given income have the same preferences and the same ability. However, we allow for $\phi \neq 0$, contrary to Mirrlees' assumption (expression 10). We start with analytical results for a simplified version of the tax designer's problem to clarify the main mechanism at work. Then, we derive numerical results for a wide range of model parameterizations.

[12] We do not claim that ours is the only plausible approach to normalizing preferences. This is a relatively simple and transparent option, and our results do not depend on the specifics of normalization so long as preferences are treated as less deserving of redistribution than ability. Indeed, the more general normative treatment of preferences made possible by using α_i weights is why we do not incorporate the normalization implied by Preference Neutrality directly into the individual utility function.

3.1 Analytical results in a simplified tax design problem

In this example we impose several simplifying assumptions on the Tax Design Problem of expressions (11) through (14) to provide the clearest picture of the effect of preferences on optimal policy. In particular, we show that a role for preferences changes optimal policy even in the first-best scenario (i.e., with no incentive effects of taxation).

First, we assume that there are only two types of individuals, a low type and a high type: $I = 2$ and $\lambda_l < \lambda_h$. With only two types, the tax system is reducible to a transfer from one type to the other. We label that transfer t, and without loss of generality assume that $c_l = y_l + t$ while $c_h = y_h - t$ so that $t > 0$ is a transfer from the high type to the low type.

Second, individuals do not respond to the tax system, instead choosing how much income to earn as if $t = 0$. For our present purposes, this assumption is not as restrictive as it may first appear. With $I = 2$, we are studying whether the high type is taxed to support the low type and how this varies with ϕ. The answers to those questions are determined by the tax designer's objective function, not the constraints it faces.

Third, we choose parameter values that make our analytical results cleaner. We set $\gamma = 1$ and $\sigma = 2$, implying logarithmic utility of consumption and quadratic disutility of labor effort.

Using expression (9) and the assumptions just stated, the tax design problem in this setting is:

Simplified Tax Design Problem

$$\max_t \left\{ \alpha_l \left[\lambda_l^{2\phi} \ln(y_l + t) - \frac{1}{2} \left(\frac{y_l}{\lambda_l^{(1-\phi)}} \right)^2 \right] + \alpha_h \left[\lambda_h^{2\phi} \ln(y_h - t) - \frac{1}{2} \left(\frac{y_h}{\lambda_h^{(1-\phi)}} \right)^2 \right] \right\}, \tag{16}$$

where each individual chooses income as if $t = 0$*, implying that for* $i \in \{l, h\}$*:*

$$y_i = \lambda_i, \tag{17}$$

and Preference Neutrality implies that for $i \in \{l, h\}$*:*

$$\alpha_i = (\lambda_i)^{-1}. \tag{18}$$

Note that condition (17) implies that the high type chooses to earn more income than the low type in this setup. The derivation of expression (18) is in the Appendix.

The first-order condition for the Simplified Tax Design Problem is:

$$t = \frac{\lambda_l^{(2\phi-1)}\lambda_h - \lambda_h^{(2\phi-1)}\lambda_l}{\lambda_h^{(2\phi-1)} + \lambda_l^{(2\phi-1)}}. \tag{19}$$

We are interested in the effect of the assumed value for ϕ on t. When $\phi = 0$, as in the conventional model, expression (19) implies $t > 0$. When all variation in earnings is attributed to preferences, so that $\phi = 1$, the optimal policy sets $t = 0$. Using (19), we can directly show the following result.

Proposition 1 *The solution to the Simplified Tax Design Problem yields the transfer t satisfying:*

$$\frac{\partial t}{\partial \phi} = \frac{2\left(\ln \lambda_l - \ln \lambda_h\right)\left(\lambda_l + \lambda_h\right)\lambda_h^{(2\phi-1)}\lambda_l^{(2\phi-1)}}{\left(\lambda_h^{(2\phi-1)} + \lambda_l^{(2\phi-1)}\right)^2} < 0. \tag{20}$$

This proposition shows that optimal redistribution, as measured by the transfer from the high-income individual to the low-income individual, decreases whenever a greater portion of income variation is attributed to preferences. By implication, any value of $\phi > 0$ will generate a lower optimal tax rate than the conventional Mirrleesian approach recommends.[13]

Why is the optimal extent of redistribution inversely related to ϕ? If $\phi = 0$, the social welfare function in (16) implies that the high type generates less disutility from earning a given income than does the low type, but both generate the same utility from a given amount of consumption. Policy will maximize social welfare if it requires the high type to earn more of society's income but leaves both types with the same level of consumption. For any $\phi > 0$, the social welfare function implies that the high type generates more utility from a given amount of consumption than the low type (the high type continues to generate less disutility from earning a given income as long as $\phi < 1$, but the difference is less). Thus, policy will maximize social welfare if it allows the high type to consume more of its income.

While this simple example has allowed us to generate a clear result on the effects of preferences on redistribution, a more comprehensive and quantitative analysis requires numerical simulations of the general tax design problem. We turn to those next.

3.2 Numerical results in a calibrated tax design problem

In this section we present numerical results for optimal policy under conventional and alternative assumptions on the role of preferences in explaining income variation, as measured by the common value of ϕ. For values in the range $\phi \in [-1, 2]$, we generate optimal values for τ, the linear income tax rate chosen by the tax designer solving the Tax Design Problem of expressions (11) through (14). The budget constraint guarantees that a higher value for τ implies a larger uniform grant g, so that redistribution unambiguously increases with τ.

Our analyses yield a clear result consistent with Proposition 1 above: when more of the variation in incomes across individuals is attributed to preferences rather than ability, optimal redistribution declines.

To perform the simulations, we specify the distribution of individual types and choose values for the model's parameters. One advantage of this paper's approach to modeling heterogeneity is that we can calibrate the distribution of individual types $\{\lambda_i\}_{i=1}^{I}$ from microdata on earnings for the United States. We assume λ_i is drawn from a lognormal distribution with parameters μ_λ and σ_λ. For a simulated vector $\{\lambda_i\}_{i=1}^{I}$, we can calculate the resulting income distribution $\{y_i\}_{i=1}^{I}$ for given tax parameters τ and g. Kotlikoff and Rapson (2006) estimate that U.S. average net marginal tax rates are in the neighborhood of 40%, and that the implied demogrant is approximately \$20,000. Therefore, using $\tau = 0.4$ and an appropriate value for g, we find the parameters of the λ distribution that minimize the sum of squared differences between our simulated income quintile thresholds and those reported by the U.S. Census Bureau in 2004. The resulting parameter estimates, using 5,000 agents (and rounded to the nearest 0.01) are $\mu_\lambda = 1.65$ and $\sigma_\lambda = 0.75$.

[13] The inequality in expression (20) holds for all assumptions on the welfare weights $\{\alpha_i\}_{i=l,h}$ so long as the weights are independent of ϕ. While Preference Neutrality seems the natural assumption to us, it is not required for the result.

Our conceptual results are not sensitive to these parameter values, but having a realistic calibration makes the magnitudes of our results easier to interpret. We consider a range of parameter values that spans most mainstream estimates: $\gamma \in \{0.5,\ 1,\ 2\}$ and $\sigma \in \{1.5,\ 3,\ 6\}$. For brevity, at times we will use a specific baseline specification, for which we assume $\gamma = 1$ and $\sigma = 3$. We choose values for the welfare weights α_i to satisfy Preference Neutrality, as described in the Appendix.

Figure 1 plots τ against ϕ, showing how optimal redistribution varies with the role attributed to preferences. The Mirrleesian benchmark is at $\phi = 0$, while the case in which all variation is attributed to preferences is at $\phi = 1$. Each subplot is a (γ, σ) pair.

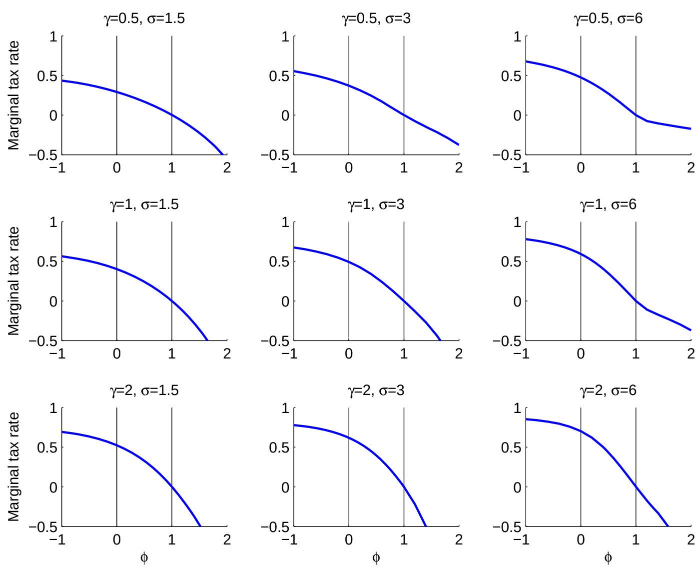

Figure 1: Optimal linear tax rate as a function of ϕ for a range of parameter values

Figure 1 shows that the optimal linear tax rate is monotonically decreasing with the importance attributed to preferences rather than ability in explaining income variation for all sets of parameter values. Note that the values of $\{\alpha_i\}_{i=1}^{I}$ cause the optimal τ for $\phi = 1$ to be zero in all cases.

Figure 1 also shows that our model clarifies the conditions under which incorporating preferences would *increase* the optimal extent of redistribution relative to the conventional approach. Optimal redistribution exceeds the conventional model's benchmark when ϕ falls below the Mirrleesian assumption of $\phi = 0$.

The key question raised by this finding is whether values of $\phi < 0$ are reasonable. It turns out that $\phi < 0$ implies a counterintuitive feature of the individual utility function, suggesting that the conventional value of $\phi = 0$ is a special and, in an important sense, extreme assumption. We formalize this in the following proposition, the proof of which can be found in the Appendix.

Proposition 2:

If $\phi < 0$, the marginal rate of substitution of consumption for leisure (i.e., the relative preference for consumption) is declining in income. That is:

$$\frac{\partial \frac{-\partial U_i(c^*,l^*)/\partial c^*}{\partial U_i(c^*,l^*)/\partial l^*}}{\partial y_i} < 0. \tag{21}$$

In words, $\phi < 0$ implies that high income earners have lower preferences for additional consumption than those who earn less if given the same starting levels of consumption and leisure. To the extent that scenario is implausible, a value of zero is a natural lower bound on ϕ.[14]

This section has shown that optimal redistribution declines when more of the variation in income is attributed to greater tastes for consumption among high earners rather than greater income-earning abilities.

This section's model captures only one way in which preferences may be thought to interact with ability, however. Thus far we have assumed all individuals with a given income have the same preferences and ability. Next, we analyze the more realistic case in which preferences may vary conditional on income.

4 Optimal policy with variation in ϕ_i

In this section, we consider variation in ϕ_i, which implies variation in preferences conditional on income. We work with the same Tax Design Problem (expressions 11 through 14) and rely on numerical simulations, given the analytical complexity of this case. We find a striking result: conditional preference heterogeneity is even more likely to imply lower optimal redistribution. Specifically, optimal redistribution is lower than in the conventional model for every one of the wide variety of distributions of $\{\phi_i\}_{i=1}^{I}$ that we consider in which we *retain* the conventional assumption that $E_i[\phi_i] = 0$.[15]

We show results for simulations in which $\phi_i \sim N(0, \sigma_\phi)$; that is, ϕ_i follows a Normal distribution centered at $E_i[\phi_i] = 0$ with variance σ_ϕ. Using alternative, asymmetric distributions for ϕ_i does not change our findings.[16] Figure 2 plots the optimal τ against σ_ϕ in this case (other parameters are at baseline levels).

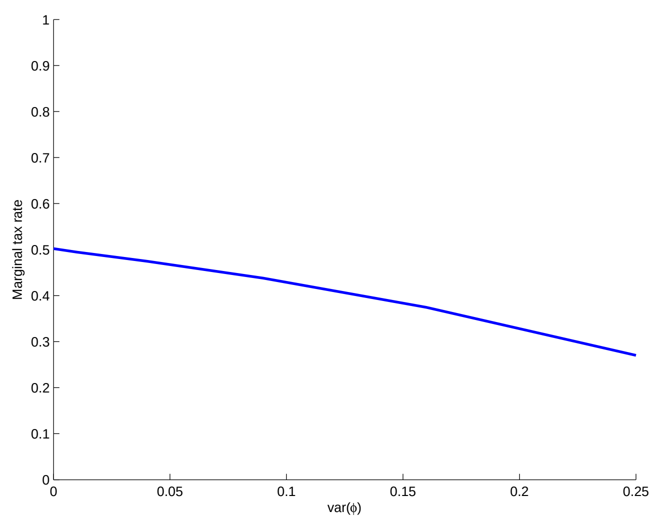

Figure 2: Optimal tax rate as a function of the variance of ϕ.

[14] Related, $\phi > 1$ implies that income-earning ability and income are negatively related.

[15] Throughout, we assume that the distributions of $\{\phi_i\}_{i=1}^{I}$ and $\{\lambda_i\}_{i=1}^{I}$ are independent. Independence means that the variation in the marginal rate of substitution of consumption for leisure is similar at each level of earnings. That is, high earners are no more or less variable in their preference for consumption relative to leisure than are low earners. Though an analysis with interdependent distributions may be of interest, we leave that to future work.

[16] For example, define the lognormally-distributed variable $X \sim \ln N(0, 0.5)$. Consider two distributions with $E_i[\phi_i] = 0$ first, $\phi = X - E[X]$, which yields a right-skewed distribution; second, $\phi = E[X] - X$, which yields a left-skewed distribution. The optimal τ for these cases are negative and 0.39, respectively, whereas τ in the conventional case is 0.50.

Figure 2 shows that the more heterogeneity of this type that one adds to the model, conditional on a given amount of income variation, the lower the optimal level of redistribution.

A related implication is shown in Figure 3, where we allow $E_i[\phi_i]$ to differ from zero. Figure 3 plots the $(\sigma_\phi, E_i[\phi_i])$ pairs that generate the same τ as in the conventional model (located at coordinates $\langle 0, \dot{0} \rangle$).

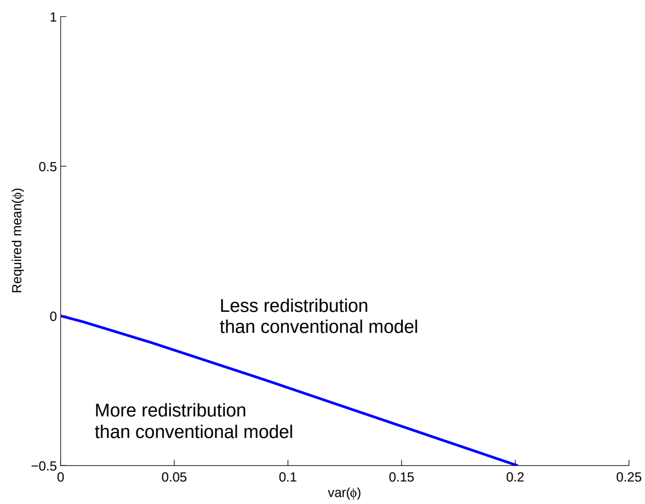

Figure 3: Pairs of $\{Var(\phi), E[\phi]\}$ yielding conventional optimal tax rate

Figure 3 shows that conditional preference heterogeneity lowers the value of $E_i[\phi_i]$ for which optimal redistribution equals the conventional model's recommendation. In other words, increased conditional preference heterogeneity is consistent with redistribution at or above the level of the conventional model only if it is offset with a lower value for $E_i[\phi_i]$, that is, only if preferences for consumption are, on average, negatively related to income earned.

These results suggest that allowing preference heterogeneity into the conventional approach reduces redistribution even if the Mirrleesian assumption is correct on average (i.e., for the average individual).

Intuitively, variation in ϕ_i means that income will be a noisier signal of ability. Redistribution will be less accurate: some high earners will have low ability and some low earners will have high ability, so a tax designer weighing the distortionary costs of redistribution against smaller benefits will choose a lower optimal level.

Related to this intuition is how different assumptions on the extent of preference heterogeneity affect the implied distributions of ability for a given distribution of earnings. Expression (8) defined the measure of income-earning ability for individual i in our model. Figure 4 shows, for the same calibrated distribution of λ_i, the distributions of ability that are implied when we vary the assumed distribution of ϕ_i. We consider the standard Mirrleesian case, in which $(E_i[\phi_i], Var[\phi_i]) = (0, 0)$, and three additional cases: $(0.2, 0.0)$, $(0.0, 0.1)$, and $(0.2, 0.1)$. The latter three cases correspond to the structure of the paper, where we first allowed for no conditional variation in preferences but increased the role for preferences on average, then allowed for conditional variation but with the conventional role for preferences on average, and finally allowed for both conditional variation and a greater role for preferences on average. The optimal τ for these cases are: 0.50, 0.43, 0.45, and 0.37.

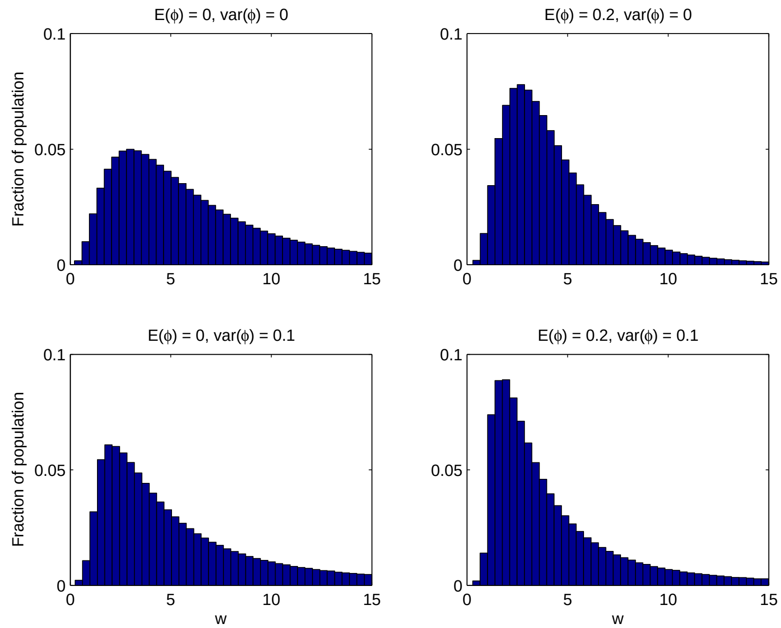

Figure 4: Ability distributions implied by different assumptions on $E[\phi_i], Var[\phi_i]$

As the figure shows, the implied distributions of ability are more compressed when the role for preferences in explaining variation in earnings is greater on average (i.e., when moving from the first to the second column of Figure 4). With less variation in ability, the tax designer gains less social welfare from redistributing, so the optimal extent of redistribution is lower. Increasing the conditional variation in preferences (i.e., moving from the first to the second row of Figure 4) may generate more or less compressed ability distributions, but it weakens the relationship between observed income and ability, making income a noisier indicator of ability and lowering optimal redistribution, as well.

In this and the preceding section we have established that the standard model occupies a special, and arguably extreme position in favor of redistribution if preferences for consumption relative to leisure vary across individuals. This conclusion may help explain the puzzle noted by Diamond and Saez (2011) that top marginal tax rates in the United States are substantially lower than conventional theory recommends. It also suggests that we might look for evidence of preference heterogeneity's effects in real-world policy.

5 Empirical evidence on preferences and policy

In this section, we translate the results of the previous sections into an empirically-testable prediction for policy: *more variation in reported preferences for consumption relative to leisure ought to be associated with less redistribution, once we control for the distribution of earnings and the extent of correlation between reported preferences and earnings.* Then, we use survey data on preferences and standard measures of redistribution to confirm that this prediction holds across both countries and U.S. states.

5.1 Deriving a testable prediction

In previous sections, we showed that optimal redistribution is inversely related to both $E[\phi_i]$ and $Var[\phi_i]$. The distribution of ϕ_i is unobservable, however, so these results are not directly testable using data on economic behavior.

It turns out that survey data on preference heterogeneity may help us to learn about the distribution of ϕ_i and, therefore, test the predictions of the theory. To see how, start with the assumption that individuals responding to survey questions on preferences over leisure and material goods are reporting information about their marginal rates of substitution (MRS) of consumption for leisure. For individual i given the allocation (c^*, l^*), the expression for the MRS is:

$$MRS_i = \frac{-\frac{\partial U_i(c^*,l^*)}{\partial c^*}}{\frac{\partial U_i(c^*,l^*)}{\partial l^*}} = \lambda_i^{(\sigma+\gamma-1)\phi_i} \frac{\left(c^*\right)^{-\gamma}}{\left(l^*\right)^{\sigma-1}}. \tag{22}$$

Without loss of generality, normalize $MRS_1 = 1$. Then, the ratio MRS_i/MRS_1 is simply the normalized preference parameter from expression (7) :

$$\theta_i = \lambda_i^{(\sigma+\gamma-1)\phi_i}. \tag{23}$$

In words, θ_i measures the (normalized) strength of each individual i's willingness to work to obtain consumption. Next, manipulate expression (23) to obtain $Var[\ln(\theta_i)] = Var[(\sigma+\gamma-1)\phi_i \ln \lambda_i]$. Using the standard expression for the variance of a product of independent[17] variables, we can write:

$$Var[\ln(\theta_i)] / (\sigma+\gamma-1)^2 = Var[\ln \lambda_i](E[\phi_i])^2 + \left[(E[\ln \lambda_i])^2 + Var(\ln \lambda_i)\right] Var(\phi_i). \tag{24}$$

Condition (24) links variation in reported preferences (θ_i) to the properties of the distribution of ϕ_i.

Suppose that a tax designer observes an increase in the variance of reported preferences: i.e., an increase in the left-hand side of expression (24). This implies some changes to $E[\phi_i]$ or $Var(\phi_i)$, given fixed values for σ, γ, and $\{\lambda_i\}_{i=1}^{I}$. What do these changes imply for optimal redistribution?

In some cases, this increased variation in reported preferences would directly imply that redistribution ought to fall. In particular, it could be that preferences play a greater role in explaining income variation on average: that is, $(E[\phi_i])^2$ rises. Or, it could be that there is an increase in the variation of preferences conditional on income: that is, greater $Var(\phi_i)$. In either case, controlling for σ and the distribution of λ_i, the tax designer knows that the optimal tax rate and extent of redistribution are lower.

A third possibility, however, complicates the relationship between reported preferences and optimal redistribution. If $Var(\phi_i)$ increases but $(E[\phi_i])^2$ decreases, the right-hand side of expression (24) may still increase (note that the two terms are multiplied by different factors) while the optimal extent of redistribution *increases* (due to the decrease in $E[\phi_i]$). This possibility suggests that the tax designer requires additional information to accurately interpret changes in reported preferences.

It turns out that a single additional observable statistic–the correlation of reported preferences and income–is sufficient to clarify the implications for optimal redistribution. To see why, note that for more variation in reported preferences to be consistent with greater optimal redistribution, $E[\phi_i]$ must decrease and $Var(\phi_i)$ must increase. In that case, preferences would be less important on average in explaining variation of incomes, and individuals with a given income would vary more in their preferences. In other

[17] See the first footnote in Section 4.

words, the correlation of income and preferences would have fallen.

Conveniently, $E[\phi_i]$ and $Var(\phi_i)$ fully determine both the variance of reported preferences (via expression 24) and the correlation of income and preferences in this model, given σ, γ, and the distribution of λ_i.[18] Any given pair $\{E[\phi_i], Var(\phi_i)\}$ also determines the optimal marginal tax rate, as in Figure 3 above. Thus, the tax designer can generate a figure such as the following to guide policy:

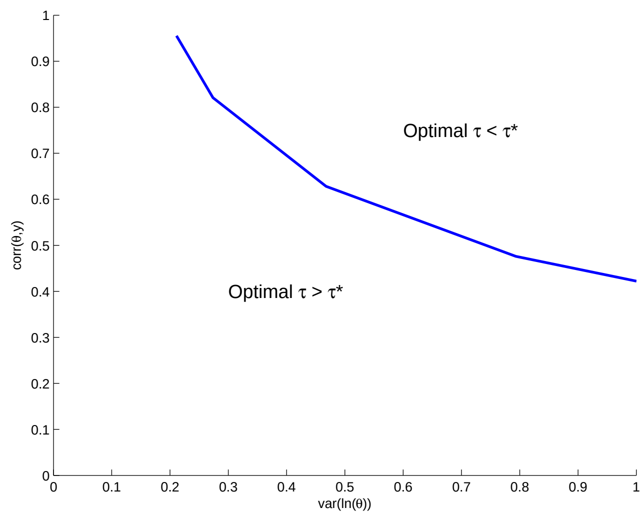

Figure 5: Optimal linear tax rate isoquant in $Var(\ln(\theta_i)), Corr(\theta_i, \lambda_i)$ space

Figure 5 plots the isoquant for an arbitrarily-chosen linear tax rate τ^* in $Var[\ln(\theta_i)], Corr(\theta_i, \lambda_i)$ space, where $Corr(\theta_i, \lambda_i)$ is the correlation between reported preferences and earnings. To the northeast of the isoquant, the optimal tax rate is less; to the southwest, it is greater.

To understand the figure, first consider a horizontal movement to the right starting at any point on the isoquant. In that case the tax designer observes an increase in the variation of reported preferences and no change in the correlation of preferences and income. This automatically implies that optimal redistribution ought to fall. Similarly, for any given variation in reported preferences, an increase in the correlation of preferences and income (a vertical movement) implies that income is a worse signal of ability, so redistribution ought to fall.

At the same time, the isoquant in Figure 5 is not a vertical line. It is possible that the variance in reported preferences could increase (a movement to the right) while the correlation of preferences and incomes could fall far enough (a movement down) so that the optimal tax rate would rise. This is the graphical version of the third possibility discussed above in the context of expression (24).

Altogether, these analyses suggest the testable prediction of the model stated at the start of this section: reported variation in preferences for consumption relative to leisure should be negatively related to the extent of redistribution, controlling for the distribution of earnings and the individual-level correlation of preferences with earnings. We turn to testing that prediction now.

[18] Income y_i and the variable λ_i are perfectly correlated, so $Corr(\theta_i, \lambda_i) = Corr(\theta_i, y_i)$.

5.2 International evidence

First, we consider cross-sectional[19] international data. Redistribution is measured by the size of social expenditures[20] as a share of GDP in 1995 as reported by the OECD. Very similar results are obtained if we use the highest marginal tax rate on personal income in 2000, the economy-wide average tax rate, or the difference between gross and net Gini coefficients as alternative measures of redistribution. Preferences are measured with responses to the World Values Survey's question C008, asked between 1995 and 2001:

> *"Which point on this scale [1 through 5] most clearly describes how much weight you place on work (including housework and schoolwork), as compared with leisure or recreation?"*
>
> *1 It´s leisure that makes life worth living, not work*
>
> *...*
>
> *5 Work is what makes life worth living, not leisure*

The variance of answers to the question serves as our measure of preference variation, formally $Var[\ln(\theta_i)]$ in expression (24) . This question is far from ideal for eliciting the marginal rate of substitution of consumption for leisure. Nevertheless, the extremes of this scale indicate fundamentally different attitudes toward the value of leisure and the question asks individuals to compare the value of leisure to the value of work, the return to which is (in this and most economic models) consumption. The distributions of responses for the countries in our sample are provided in the Appendix.

To get a sense for how the data correspond to the model's prediction, Figure 6 plots social transfers as a share of GDP against the variance of responses for fifteen OECD country observations.

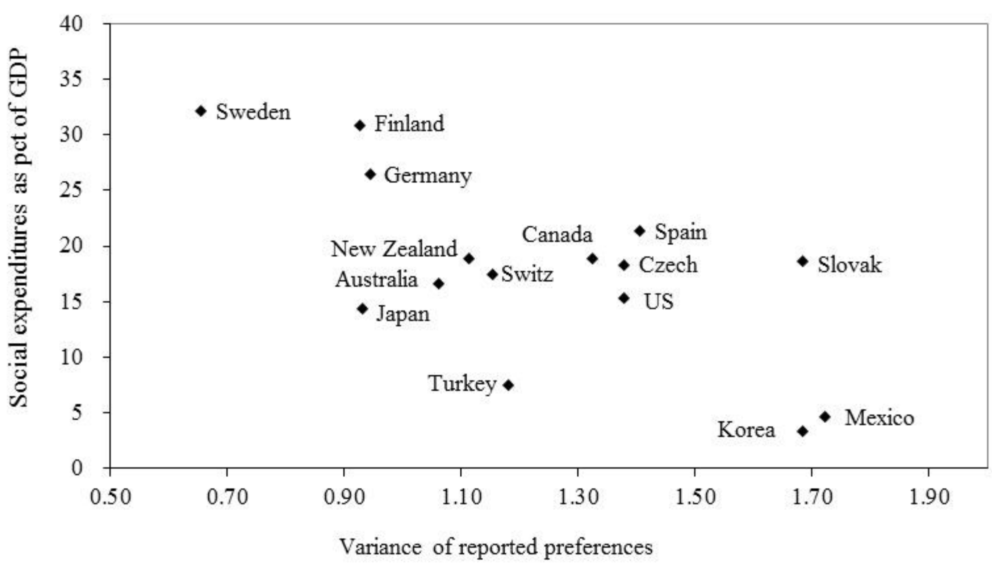

Figure 6: Redistribution and preference variation in 15 OECD countries

[19] Panel analysis would be desirable, but the survey data we use to measure preferences is available over at most a ten-year horizon. We believe this is too narrow a window over which to expect either meaningful changes in preference variation or a response to any such changes in policy, so we leave the analysis of panel data for future research.

[20] Social expenditures include the following programs, according to the OECD definition: "Old age, Survivors, Incapacity-related benefits, Health, Family, Active labor market programmes, Unemployment, Housing, and Other social policy areas." See www.oecd.org/els/social/expenditure.

A negative relationship between these variables is apparent in Figure 6. That is, countries with more variation in preferences for consumption relative to leisure appear to have less redistributive policies. Of course, Figure 6 is not convincing evidence of a statistically meaningful relationship. To more carefully study the relationship between these variables, we want to control for some important additional factors suggested by the analysis above. The results of a simple OLS regression with several such controls is shown in Table 1.

Table 1: Social transfers as a share of GDP

	(1)	(2)
Variance of reported preferences	-18.27**	-17.92**
	(5.35)	(7.64)
Log of GDP per capita		-3.67
		(6.88)
Gross Gini coefficient		-0.55*
		(0.29)
Correlation of preferences and earnings		-61.80
		(127.97)
Ethnolinguistic fractionalization		7.74
		(9.04)
Mean of reported preferences		-11.51
		(7.95)
Observations	15	15
Adjusted R-squared	0.43	0.50

Notes: Each column shows an OLS regression in which the dependent variable is social transfers as a share of GDP. The sample is OECD countries with sufficient data. Standard errors are shown in parentheses. **p<0.05; *p<0.10

The univariate regression (1) estimates the best-fit line for Figure 6 to give a baseline for comparison. Regression (2) more carefully tests the predicted relationship between preference heterogeneity and redistribution derived at the start of this section. That prediction is conditional on two things: the distribution of λ_i (laissez-faire income[21]) and the correlation of preferences to earnings. To account for the former, we control for the level and spread of pretax income with the log of GDP per capita and the (gross) Gini coefficient[22] as reported in Solt (2008-9). To account for the latter, we calculate correlations between individuals' reported preferences and income, where income is given by individuals' reports to the World Values Survey of

[21] We do not observe laissez-faire income, and instead use the distribution of pre-tax income. Given progressive taxation, the distribution of gross pre-tax income is likely to be compressed relative to the distribution of λ_i. The extent of the omitted gap between laissez-faire and pre-tax inequality will be positively correlated to redistribution. This omitted gap may bias our estimated coefficient of interest toward zero. To see why, suppose that we accurately measure the level of preference variation. But, we underestimate of the level of income variation, and the gap is greater in more redistributive countries. Then, we attribute too great a share of income variation in redistributive countries to preferences. In other words, we are biased toward finding more preference variation as a share of total variation in redistributive countries than is true in reality, pushing the coefficient on the variance of reported preferences toward zero. Another way to see this is to note that, given a distribution of ability, the extent of variation in preferences is positively related to the spread of laissez-faire income. Thus, the coefficient on the variance of preferences may absorb some of the positive correlation between the omitted income inequality and redistribution.

[22] We use the 1990 level of the Gini coefficient to avoid a concern that contemporaneous policy could affect the extent of pre-tax inequality. Similar results are obtained using 1995 or 2000 Gini coefficients, instead, though the significance level of the coefficient on the variance of reported preferences drops to the ten percent level.

their income tercile. In addition to these controls, we control for two other variables.[23] First, an alternative potential explanation for our results is that the extent of ethnic heterogeneity is driving both variance in preferences and the level of redistribution, so we control for a measure of ethnolinguistic fractionalization taken from Alesina et al. (2003). Second, we include the mean reported preference in each country to help rule out concerns with scaling.

The results confirm the model's central prediction, with heterogenous preferences statistically significantly predicting less redistribution. The coefficient on the variance of preferences changes little when we add control variables, further supporting the robustness of the relationship. Similar results hold if we use any of the alternative measures of redistribution as the explanatory variable.

Furthermore, the importance of preference heterogeneity appears to be sizeable. If we estimate regression (2) excluding it, the adjusted R-squared falls from the 0.50 reported in Table 1 to 0.25. In other words, preference heterogeneity explains one-third of the residual variance from a regression predicting the extent of redistribution across OECD countries with the other independent variables shown in Table 1.

A natural concern with this analysis is reverse causality. That is, more redistributive policy may generate more uniform labor effort across individuals and, if the survey question is eliciting marginal values of leisure (rather than the marginal rates of substitution from a common starting point that we assume in the theory), one may worry whether this would generate more uniform preferences. In fact, the bias is likely to go in the other direction. To see why, consider two individuals with the same income-earning ability but different preferences for consumption relative to leisure. If they answer the survey assuming a common allocation of consumption and leisure, we will observe the true preference difference in their responses. On the other hand, if they answer assuming a realized allocation of consumption and leisure rather than the abstract common allocation, the high-preference individual will have chosen to earn more consumption and take less leisure. For that individual, the marginal value of consumption will fall relative to leisure, so his *reported* preference for consumption will fall and we will observe a smaller difference in preferences between the two individuals. In other words, the distribution of reported preferences will be compressed relative to the distribution of true preferences if surveys are answered "on the margin". Importantly, the extent of this compression will be greatest in countries with the *least* redistributive systems, as individuals are then encouraged to earn nearer their preferred positions. Thus, reported preference variation is likely to be (misleadingly) smaller where redistribution is also smaller, inducing a positive bias to the estimated coefficient on the variance of preferences in our results above.

Our findings can be interpreted as consistent with two prominent explanations for differences in redistribution across countries. First, a large literature has argued that an individual taxpayer may be more willing to support redistribution if his country's population is more homogenous (see Desmet et al, 2009, for example). The explanation for such findings is usually that individuals have greater affinity for those like them. The results of this paper provide evidence of a specific channel, preferences over consumption relative to leisure, through which homogeneity may generate these greater feelings of camaraderie and kinship. Second, Alesina and Angeletos (2005) show that countries with less redistributive policies are those in which most individuals believe effort, rather than luck, is the main determinant of personal income. That important finding leaves open the question of where such beliefs come from. Our results can help, as they imply that in countries where more of the variation in incomes is due to heterogeneous preferences, optimal taxes are less redistributive and a larger share of high earners are willing to exert substantial effort.

[23] We have also added the variance in answers to a "placebo" question from the same survey to check for whether some countries simply give more variable answers than others. The results are essentially unchanged from those in the Table.

5.3 Evidence from U.S. states

Next, we test for whether the results across OECD countries hold at a subnational level inside the United States. Our preferred measure of redistribution is that of Feldstein and Wrobel (1998), who calculate the difference between the (statutory) average state income tax rates at the $100,000 and $10,000 income levels in 1989.[24] For preferences, we use responses to a question on the importance of material possessions from the General Social Survey administered in the United States in 1993.[25] Question 477E is:

> *"I'm going to read you a list of some things that different people value. Some people say these things are very important to them. Other people say they are not so important. Please tell me how important each things is to you personally, using the responses on this card (HAND CARD TO RESPONDENT). How about having nice things? Is it one of the most important values you hold, very important, somewhat important, not too important, or not at all important?*

1 One of the most important

...

5 Not at all important

Figure 7 plots redistribution against the variance of responses to this question.

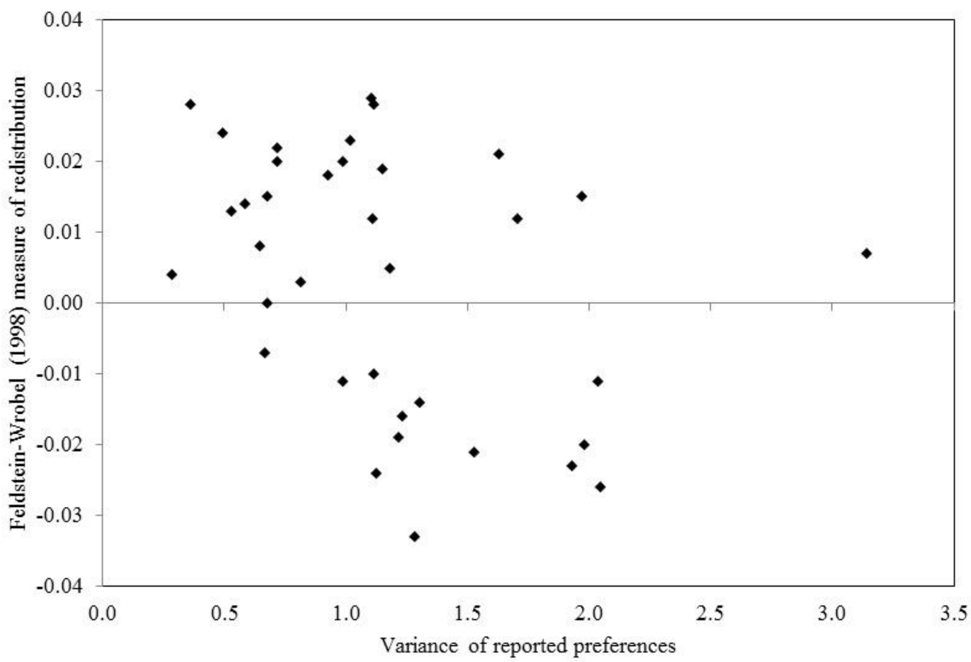

Figure 7: Redistribution and preference variation in 36 U.S. states

As did Figure 6 for the international evidence, Figure 7 displays a negative relationship between the variation in reported preferences and redistribution across U.S. states. We explore this relationship further in Table 2, which shows the results in a similar format as Table 1.

[24] The results are similar if we use the highest state marginal tax rate on personal income in 1993 as catalogued by the Tax Foundation (1994), instead.

[25] The geographic identification codes of the GSS are closely protected to ensure confidentiality. They are used in these analyses but cannot be shared by the author. To obtain the data, contact the National Opinion Research Center at www.norc.org.

Table 2: Feldstein-Wrobel (1998) measure of redistribution

	(1)	(2)
Variance of reported preferences	-0.011**	-0.014**
	(0.005)	(0.006)
Log of per capita income		-0.030
		(0.020)
Top decile share of gross income		-0.065
		(0.171)
Correlation of preferences and earnings		n/a
Fraction of state population "white"		-0.0005*
		0.0003
Mean of reported preferences		0.014
		(0.014)
Observations	36	36
Adjusted R-squared	0.11	0.18

Notes: Each column shows an OLS regression in which the dependent variable is state-level redistribution as measured in Feldstein and Wrobel (1998)..
The sample is U.S. states with sufficient data.
Standard errors are shown in parentheses. **p<0.05; *p<0.10

The univariate regression (1) again shows the best-fit line for the scatterplot as a basis for comparison. Regression (2) tests the prediction of the model. As with the same regression in Table 1, it controls for the pre-tax income distribution with measures of per capita income and income inequality, the latter measured with the share of the state's gross income reported to the IRS by the top decile of the population (see Frank, 2009; Gini coefficients at the state level were not available).[26] We do not have the individual data required to compute the correlation between preferences and earnings, so Regression (2) cannot control for that in this case (it was insignificant in the results shown in Table 1). We do control for a measure of the racial diversity within each state and the mean answer to the preferences question, as we did with the international data.

The results from regression (2) show that the impression in Figure 7 is robust: the negative coefficient on the variance of preferences changes little when we add these controls, and its significance actually increases. Moreover, we can show that preference heterogeneity explains more than ten percent of the (large) residual variation from a regression that omits the variance of reported preferences from regression (2). The U.S. state data thereby provide a second empirical example of redistributive policy consistent with the predictions of the generalized model developed in this paper.

[26] As a further test, we cluster states into either four or nine groups based on their levels of per capita income and top decile income shares. We then re-run the regressions with dummy variables for each "matched" group, relying for identification on variation in the heterogeneity of preferences within clusters. The coefficient on the variance of reported preferences in the case of four clusters is -0.016 with a (clustered) standard error of 0.008, giving a p-value of 0.13. A similar analysis across countries is infeasible because of the limited sample size.

6 Conclusion

The argument that differences in preferences, not merely ability, play a role in driving the variation in income across individuals has a long history in critiques of redistributive taxation. Nevertheless, preferences are assumed to play no such role in the leading modern model of optimal taxation.

This paper generalizes the conventional optimal tax model to include these preferences in a tractable and normatively natural way, nevertheless staying close to the conventional setup. We then derive, through both analysis and numerical simulations, novel and clear implications for how optimal redistribution in this more general model relates to conventional results. Finally, we show a new empirical finding that suggests these implications are consistent with evidence on real-world policy and reported preferences.

We find that attributing a portion of the observed variation in incomes to preferences rather than ability reduces the optimal extent of redistribution for the most plausible specifications of the model. The conventional Mirrleesian approach is a special and, arguably, extreme case. If preference heterogeneity is part of what drives differences in individuals' effort and earnings, this analysis suggests that the standard approach overstates the optimal extent of redistribution. This may help to explain (in part) a prominent puzzle in optimal tax research, noted by Diamond and Saez (2011), that existing marginal tax rates in the United States peak at a level substantially below what the conventional model recommends.

We analyze cross-country and cross-U.S.-state data on preferences and policy and show that, controlling for a variety of factors, regions with higher variance in preferences have less redistributive policies. This evidence is consistent with a direct prediction of our model.

Our findings suggest that this paper's generalized normative optimal tax model may be a better guide to policy advice than the conventional one. It also provides a novel explanation for differences in redistribution across countries and limits to redistribution within countries. A proper understanding of the role of preference heterogeneity improves our ability not only to design taxation but to understand existing tax policies.

References

[1] Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, (2003). "Fractionalization," *Journal of Economic Growth* 8.

[2] Alesina, Alberto and George-Marios Angeletos (2005). "Fairness and Redistribution," *American Economic Review*, 95(4), (September).

[3] Beaudry, Paul, Charles Blackorby, and Dezso Szalay (2009). "Taxes and Employment Subsidies in Optimal Redistribution Programs," *American Economic Review*, 99(1), pp. 216-242.

[4] Boadway, Robin, Maurice Marchand, Pierre Pestieau, and Maria del Mar Racionero (2002). "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 4(4), pp. 475-498.

[5] Caplan, Bryan (2011). Blog posting. http://econlog.econlib.org/archives/2012/01/kahneman_greed.html

[6] Chone, Philippe and Guy Laroque (2010). "Negative Marginal Tax Rates and Heterogeneity", *American Economic Review 100,* pp. 2532-2547.

[7] Diamond, Peter and Emmanuel Saez, (2011). "The Case for a Progressive Tax: From Basic Research to Policy Recommendations" forthcoming in *Journal of Economic Perspectives.*

[8] Diamond, Peter and Johannes Spinnewijn (2011) "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy* 3(4), November.

[9] Farhi, Emmanuel and Iván Werning (2011). "Insurance and Taxation over the Life Cycle," Working Paper, November.

[10] Fleurbaey, Marc and Francois Maniquet (2004). "Compensation and Responsibility," working paper from 2004 used. Forthcoming in K.J. Arrow, A.K. Sen and K. Suzumura (eds) *Handbook of Social Choice and Welfare.*

[11] Fleurbaey, Marc and Francois Maniquet (2006). "Fair Income Tax," *Review of Economic Studies*, 73, pp. 55-83.

[12] Frank, M. W. 2009 "Inequality and Growth in the United States: Evidence from a New State-Level Panel of Income Inequality Measure" *Economic Inquiry*, 47(1), pp. 55-68.

[13] Friedman, Milton (1962). *Capitalism and Freedom.* Chicago: University of Chicago Press.

[14] Golosov, Mikhael, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2011). "Preference Heterogeneity and Optimal Capital Income Taxation," *NBER Working Paper 16619.*

[15] Harsanyi, John C. (1953) "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy*, 61(5), (October), pp. 434-435.

[16] Harsanyi, John C. (1955). "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63(4), (August), pp. 309-321.

[17] Hellwig, Martin (1986). "The Optimal Linear Income Tax Revisited," *Journal of Public Economics* 31(2) pages 163–179.

[18] Judd, Kenneth and Che-Lin Su (2006). "Optimal Income Taxation with Multidimensional Taxpayer Types," Working Paper, earliest available draft April, 2006; more recent draft November 22, 2006.

[19] Kahneman, Daniel (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus, and Giroux.

[20] Kaplow, Louis (2008). "Optimal Policy with Heterogeneous Preferences," *The B.E. Journal of Economic Analysis & Policy*, 8(1) (Advances), Article 40.

[21] Kocherlakota, Narayana and Christopher Phelan (2009). "On the robustness of laissez-faire," *Journal of Economic Theory* 144, pp. 2372-2387

[22] Kotlikoff, Laurence and David Rapson (2009) "Does it Pay, at the Margin, to Work and Save?", NBER working paper 12533.

[23] Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan, (2009). "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), pages 147-174.

[24] Mirrlees, James A. (1971). "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies.*

[25] Mirrlees, James A. (1986) "The Theory of Optimal Taxation", in *Handbook of Mathematical Economics,* edited by K.J. Arrow and M.D. Intrilligator, vol. III, ch. 24, 1198-1249.

[26] Nozick, Robert (1974). *Anarchy, State, and Utopia.* New York: Basic Books.

[27] Pigou, Arthur C.(1928). *A Study in Public Finance*. Macmillan. (page 58, 1960 reprint).

[28] Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies,* 68.

[29] Saez, Emmanuel (2002). "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics.*

[30] Sandmo, Agnar (1993). "Optimal Redistribution when Tastes Differ," *Finanzarchiv* 50(2).

[31] Sheshinksi, Eytan (1972). "The Optimal Linear Income-Tax," *Review of Economic Studies* 39(3).

[32] Solt, Frederick (2008-09). "The Standardized World Income Inequality Database", http://hdl.handle.net/1902.1/11992 V3

[33] Tax Foundation (1994). "State Tax Rates and Collections in 1993," *Special Report* Number 27

[34] Tarkiainen, Ritva and Matti Tuomala (2004). "On Optimal Income Taxation with Heterogeneous Work Preferences," Tampere Working Economic Papers.

[35] Tuomala, Matti (1990). *Optimal Income Tax and Redistribution*, Oxford University Press.

[36] Vickrey, William (1945). "Measuring Marginal Utility by Reactions to Risk," *Econometrica*, 13(4), (October), pp. 319-333.

Appendix

Deriving welfare weights that yield Preference Neutrality

Welfare weights can be chosen so that differences in preferences justify no redistribution in the Simplified Tax Design Problem, i.e., to satisfy Preference Neutrality. With unrestricted α_i the expression for optimal t in the Simplified Tax Design Problem with $\phi = 1$ is

$$t = \frac{\alpha_l \lambda_l^2 \lambda_h - \alpha_h \lambda_h^2 \lambda_l}{\alpha_h \lambda_h^2 + \alpha_l \lambda_l^2}. \tag{25}$$

To obtain optimal $t = 0$, the tax designer sets

$$\alpha_i = 1/\lambda_i \tag{26}$$

for $i = \{l, h\}$.

A generalized version of this adjustment to the the Simplified Tax Design Problem applies to the general Tax Design Problem. In particular, the expression:

$$\alpha_i = \lambda_i^{-\mu(\sigma+\gamma-1)\phi} \tag{27}$$

allows the tax designer to adjust the normative treatment of preferences by setting the new parameter $\mu \in [0, 1]$. In the Simplified Tax Design Problem, $\mu = 1/2$ yielded an optimal policy of no redistribution when all heterogeneity was assigned to preferences ($\phi = 1$).

For the general Tax Design Problem and the numerical simulations in the paper, we are looking for values of α_i so that no redistribution is justified when $\phi_i = 1$ for all i. Because a policy of no redistribution is incentive compatible, we can simplify the Tax Design Problem to choosing a vector of taxes subject to feasibility only. That social planner's problem is:

$$\max_{\{t\}_{i=1}^{I}} \sum_{i=1}^{I} \alpha_i \left[\lambda_i^{(\sigma+\gamma-1)\phi_i} \frac{(y_i - t_i)^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma} \left(\frac{y_i}{\lambda_i^{(\sigma+\gamma-1)(1-\phi_i)/\sigma}} \right)^{\sigma} \right], \tag{28}$$

subject to feasibility

$$\sum_{i=1}^{I} (t_i) \geq 0$$

The planner's first-order conditions imply:

$$\frac{\alpha_i}{\alpha_j} \left(\frac{\lambda_i}{\lambda_j} \right)^{(\sigma+\gamma-1)\phi_i} \left(\frac{c_i}{c_j} \right)^{-\gamma} = 1.$$

In the laissez faire, $c_i = y_i$ and each agent sets $y_i = \lambda_i$ as discussed in the setup of the model. For the optimal allocation to match the laissez-faire when $\phi = 1$, this implies:

$$\frac{\alpha_i}{\alpha_j} \left(\frac{\lambda_i}{\lambda_j} \right)^{(\sigma-1)} = 1.$$

Thus, choosing

$$\alpha_i = \lambda_i^{(1-\sigma)}$$

neutralizes preferences' effects on redistribution. Note that any multiple of this vector would work as well.

Generally, the social planner's maximization can be rewritten as:

$$\max_t \sum_{i=1}^{n} \lambda_i^{(\sigma+\gamma-1)\phi(1-\mu)} \frac{{c_i}^{1-\gamma}-1}{1-\gamma} - \frac{1}{\sigma\lambda_i^{(\sigma+\gamma-1)\phi\mu}} \left(\frac{y_i^\star}{\lambda_i^{(\sigma+\gamma-1)(1-\phi)/\sigma}} \right)^\sigma ,$$

where we can implement Preference Neutrality as defined in the text by choosing

$$\mu = \frac{(\sigma-1)}{(\sigma+\gamma-1)}.$$

In general, μ can be varied to adjust the cardinalization of preferences in the social welfare function.

Proof of Proposition 2

To derive result (21), start with expression (9) to derive the marginal rate of substitution for individual i starting at the allocation (c^*, l^*) :

$$MRS_i = \frac{-\frac{\partial U_i(c^*,l^*)}{\partial c^*}}{\frac{\partial U_i(c^*,l^*)}{\partial l^*}} = \lambda_i^{(\sigma+\gamma-1)\phi} \frac{\left(c^*\right)^{-\gamma}}{\left(l^*\right)^{\sigma-1}}. \tag{29}$$

Differentiating with respect to λ_i gives:

$$\frac{\partial MRS_i}{\partial \lambda_i} = (\sigma+\gamma-1)\phi\lambda_i^{(\sigma+\gamma-1)\phi-1} \frac{(c^*)^{-\gamma}}{(l^*)^{\sigma-1}}. \tag{30}$$

Note that the conventional assumption, $\phi = 0$, applied to expression (30) implies:

$$\phi = 0 \Rightarrow \frac{\partial MRS_i}{\partial \lambda_i} = 0.$$

In words, the conventional assumption is that all individuals place the same marginal value on consumption relative to leisure when starting at a given allocation of them. In contrast,

$$\phi < 0 \Rightarrow \frac{\partial MRS_i}{\partial \lambda_i} < 0, \tag{31}$$

so that $\phi < 0$ implies that individuals with higher λ_i are less willing to work in order to increase their consumption than are individuals with lower λ_i. To connect this result to income earned, use expression (14) , which shows that the value of ϕ is immaterial to individual i's choice of earnings, given a tax system and a value for λ_i. Applying the implicit function theorem to that expression, we can show that

$$\frac{\partial y_i}{\partial \lambda_i} = \frac{(\sigma+\gamma-1)(1-\tau)\lambda_i^{(\sigma+\gamma-1)-1}(g+(1-\tau)y_i)^{-\gamma}}{\gamma(1-\tau)(1-\tau)\lambda_i^{(\sigma+\gamma-1)}(g+(1-\tau)y_i)^{-\gamma-1}+(\sigma-1)(y_i)^{\sigma-2}} > 0 \text{ if } \tau < 1. \tag{32}$$

Result (32) implies that an individual with higher λ_i will choose to earn more income than an individual with lower λ_i. Combining results (31) and (32) yields the relationship described in the proposition.

Responses to World Values Survey question on preferences

Here, we show the distribution of responses to the question cited in the text for the fifteen OECD countries in our sample.

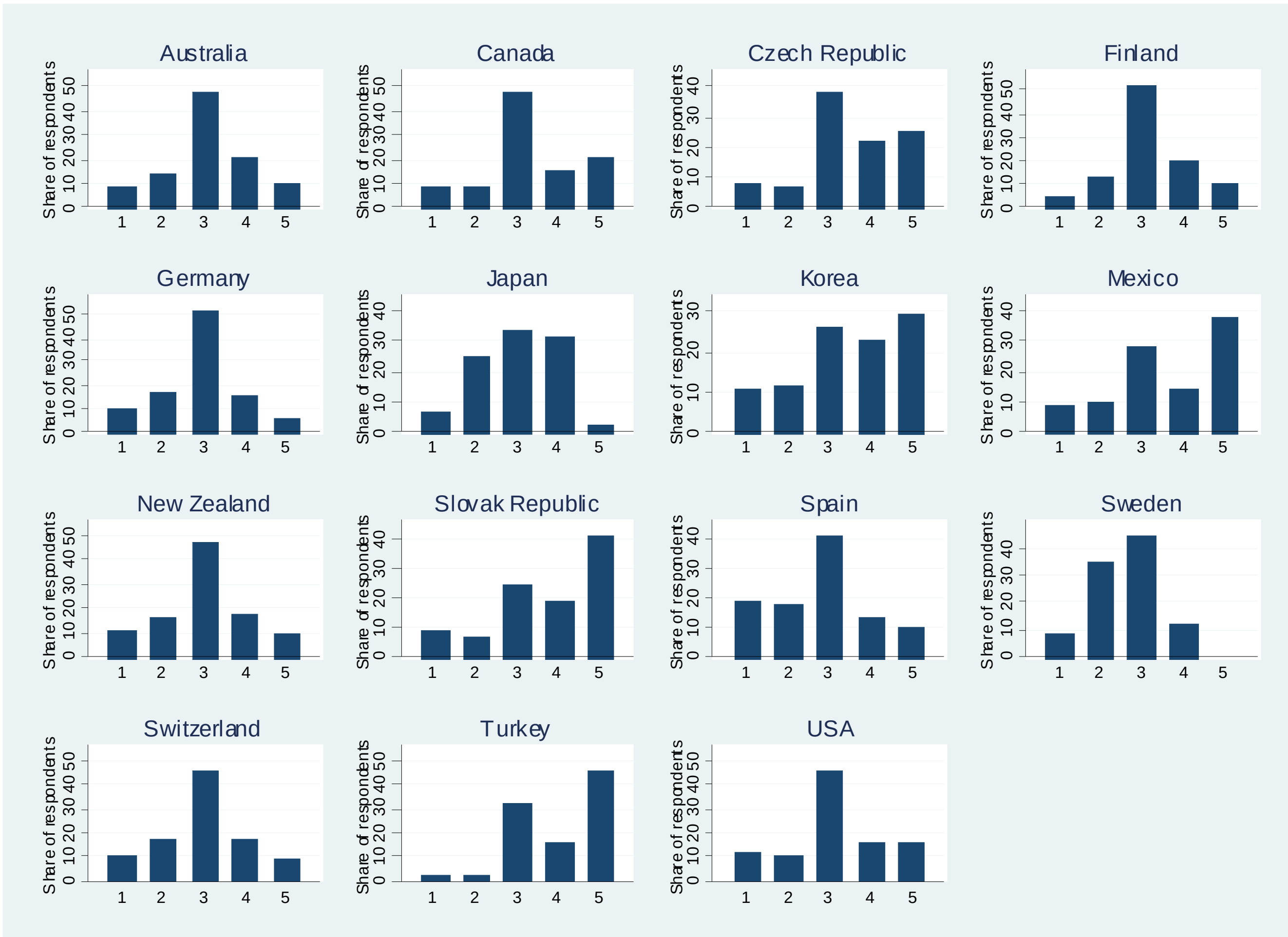

Figure A1: Distribution of responses within each country

Most of these distributions are centered away from the extreme values, suggesting that truncation is not a major concern. Even those with substantial mass at the extremes, i.e., Korea and Mexico, have sizable proportions of respondents at all values.