NBER WORKING PAPER SERIES

DE GUSTIBUS NON EST TAXANDUM:
HETEROGENEITY IN PREFERENCES AND OPTIMAL 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, Dan Shoag, Aleh Tsyvinski, Glen Weyl, Danny Yagan, anonymous referees 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: Heterogeneity in Preferences and Optimal Redistribution
Benjamin B. Lockwood and Matthew C. Weinzierl
NBER Working Paper No. 17784
January 2012, Revised April 2014
JEL No. H2,H21

ABSTRACT

The prominent but unproven intuition that preference heterogeneity reduces re-distribution in a standard optimal tax model is shown to hold under the plausible condition that the distribution of preferences for consumption relative to leisure rises, in terms of first-order stochastic dominance, with income. Given mainstream functional form assumptions on utility and the distributions of ability and preferences, a simple statistic for the effect of preference heterogeneity on marginal tax rates is derived. Numerical simulations and suggestive empirical evidence demonstrate the link between this potentially measurable statistic and the quantitative implications of preference heterogeneity for policy.

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

In the early years of modern optimal tax research, theorists assumed all individuals have the same preferences over consumption and leisure. James A. Mirrlees's (1971) second simplifying assumption was that "Differences in tastes...are ignored. These raise rather different kinds of problems, and it is natural to assume them away." This simplification freed Mirrlees to assume that the only way in which people differ is in their ability to earn income.[1] His powerful approach–complete with its assumption of preference homogeneity–now dominates theoretical work on tax design.

The omission of this form of preference heterogeneity has cast a shadow over optimal tax research from the beginning, however, both because differences in these preferences are an evident feature of reality and because omitting them is far from innocuous in Mirrlees's setting. Kahneman (2011) reports that such preference differences are widespread among young adults and correlate with outcomes later in life. Data shown in Section 3 of this paper, from the World Values Survey and General Social Survey, reveal that respondents report a wide range of views toward the value of material possessions. More anecdotally, people choose a wide range of consumption-leisure bundles, even conditional on apparent budget constraints. Importantly, these preferences over consumption and leisure are observationally equivalent to the income-earning ability at the heart of the Mirrlees approach. That is, an individual may earn a low income, and respond to taxes the way he does, either because he has low ability or because he has a weak relative preference for consumption. As has long been recognized, for instance in the formal literature by Agnar Sandmo (1993), the possibility of these two cases presents a challenge for optimal tax design. If an individual is to be compensated for having low ability but held responsible for his preferences, in the terms of Marc Fleurbaey and Francois Maniquet (2004), society will want to redistribute income to a low earner in the first case but not in the second.

Complicating matters is that the effect of preference heterogeneity on optimal redistribu-

[1] Mirrlees was not the first to adopt this simplification. Arthur 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."

tion is ambiguous. While critics of redistribution have long stressed that income may signal preferences rather than ability[2], in principle adding preference heterogeneity to the model may *increase* optimal redistribution. Intuitively, if preferences for consumption relative to leisure are lower among those with high incomes, attributing income variation to ability alone will understate the income-earning abilities of high earners and imply an optimal extent of redistribution that is too small. This theoretical ambiguity has been left unresolved in general settings by previous analyses.[3] As Louis Kaplow (2008) notes, "This difficulty is not merely technical; many have suspected that the nature of the optimum may change in important ways...Although such conjectures have been informally expressed, the issue has received little sustained attention."

In this paper, we derive two novel results that clarify preference heterogeneity's effects on the optimal extent of redistribution.

First, we generalize a standard optimal tax model to compare the extent of redistribution under an optimal policy in which preference heterogeneity is allowed to exist to that under the Mirrleesian benchmark policy in which all heterogeneity is assumed to be in ability. Our model fully nests the conventional model. We show that the prominent but unproven intuition that heterogeneity in preferences over consumption and leisure lowers optimal redistribution is incomplete but correct under a specific, plausible condition. In particular, if the distribution of the relative preference for consumption over leisure rises with income (in

[2]See Robert Nozick (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?" Or, Milton Friedman (1962), "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."

[3]Mirrlees (1976, 1986) addressed the general case but obtained inconclusive results. Some prior work adopts specialized settings, such as Sandmo's (1993) insightful analysis with only preference (not ability) heterogeneity; Robin Boadway, Maurice Marchand, Pierre Pestieau, and Maria del Mar Racionero's (2002) results with two preference types, two ability levels, and quasilinear utility; and Fleurbaey and Maniquet's (2006) analysis with a specific normative approach. Other work has focused on numerical simulations, such as Ritva Tarkiainen and Matti Tuomala (2004) or Kenneth L. Judd and Che-Lin Su (2006), who explain the computational complexities associated with multiple dimensions of heterogeneity. Two other recent papers focus on related but somewhat different questions. Narayana Kocherlakota and Christopher Phelan (2009) focus on the policy implications of uncertainty over 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. Paul Beaudry, Charles Blackorby, and Dezso 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.

terms of first-order stochastic dominance), then optimal marginal tax rates are lower at all incomes, and redistribution to the lowest earner is less, than in the Mirrleesian case. We show this result analytically for the case of quasilinear utility studied in Diamond (1998), and we show through numerical simulations that it holds for more general utility functions with income effects. As a consequence, the conventional Mirrleesian assumption of preference homogeneity ought not to be seen as neutral with respect to determining the optimal extent of redistribution.

Second, we derive a simple statistic for the effect of heterogeneity in preferences on optimal marginal tax rates assuming plausible functional forms for ability and the distributions of ability and preferences. We demonstrate the link between this statistic and the quantitative implications of preference heterogeneity for optimal policy using numerical simulations calibrated to the U.S. economy. We also generate empirical estimates of this statistic for OECD countries and U.S. states and use them to show suggestive evidence that existing policy variation appears to be consistent with our theoretical findings. Though this simple statistic is not observable through conventional economic variables, our findings suggest it is a valuable target for future empirical work.

We obtain our novel analytical results by combining two recent innovations in the literature with a third innovation of our own. First, in a setting with a continuum of agents and standard utility functions, Philippe Choné and Guy Laroque (2010) show how heterogeneity in the opportunity cost of work can justify negative marginal tax rates at low incomes. They achieve this important finding in part by collapsing multiple dimensions of heterogeneity into a single composite characteristic that determines behavior. This technique, similar to that used by Craig Brett and John Weymark (2003), avoids the technical obstacle of multi-dimensional screening that had limited much of the prior work on preference heterogeneity.[4] We adopt an approach akin to Choné and Laroque's in order to reduce to a single dimension the heterogeneity driving individual behavior. While this technique cannot help with all conceivable dimensions of heterogeneity,[5] it is well-suited for the preferences over con-

[4]Casey Rothschild and Florian Scheuer (2013) use a different method to avoid the technical problems with multi-dimensional income-earning ability.

[5]For example, time discounting as in Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2013) or Peter Diamond and Johannes Spinnewijn (2011).

sumption and leisure studied in this paper. Second, we adopt the moral reasoning behind the "second fairness requirement" in the prominent work of Marc Fleurbaey and Francois Maniquet (2006), which states that "the laisser-faire (this is, the absence of redistribution) should be the social optimum in the hypothetical case when all agents have equal earning abilities" even if they have different preferences.[6] In other words, we adopt the normative perspective that preferences over consumption and leisure do not justify redistribution by themselves. Though specific, this interpretation is the natural one if preferences are thought of as tastes as opposed to, for example, needs (see Kaplow 2008 for a discussion). Third, and crucially, we introduce the technique of studying how optimal policy changes when a given distribution of income is attributed to more than one source of heterogeneity, rather than how optimal policy changes when ability is augmented with additional sources of heterogeneity that change the distribution of income. This shift makes possible our progress over prior results. It has the additional virtue of formulating the problem in a way resembling that faced by policy makers, who must decide the appropriate extent of redistribution in the face of an observable income distribution.

The paper proceeds as follows. Section 1 presents the generalized Mirrleesian optimal tax model that incorporates preference heterogeneity. Section 2 derives the analytical results on optimal redistribution with quasilinear utility. Section 3 derives a simple statistic for marginal tax rates for a familiar set of functional form assumptions and uses it to quantify the potential impact of preference heterogeneity on optimal policy through both numerical simulations and empirical evidence. The numerical simulations also demonstrate the robustness of the analytical findings from Section 2 to the use of a utility function that includes income effects. Section 4 provides a concluding discussion.

[6] Fleurbaey and Maniquet (2006) 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. 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. Our analysis can be seen as a complement to theirs, studying the same type of preference heterogeneity in a setting closer to the more conventional Mirrleesian approach.

1 A generalized model with preference heterogeneity

Our model closely follows (and fully nests) the conventional Mirrleesian setup in which a social planner designs taxes to maximize social welfare and a population of heterogeneous individuals choose labor effort, taking the tax system as given, to maximize their individual utilities.

1.1 Individuals

A set of individuals of measure one, indexed by i, differ in both their abilities to convert labor effort into earned income and, in a departure from the standard approach, their preferences for consumption relative to leisure. Individual i derives utility from consumption and leisure (equivalently, disutility from labor effort) according to the following function:

$$U_i(c_i, l_i) = U\left(c_i, \frac{l_i}{\theta_i}\right), \tag{1}$$

where $c_i \geq 0$ denotes consumption, $l_i \geq 0$ denotes labor effort, and $\theta_i > 0$ denotes preferences–all for individual i. We assume the standard conditions on the utility function $U : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$, namely $U_c > 0$, $U_{cc} < 0$, $U_l < 0$, $U_{ll} < 0$. As in the conventional model, labor effort for individual i is calculated as $l_i = y_i/w_i$, where $y_i \geq 0$ denotes earned income and $w_i \geq 0$ denotes income-earning ability.[7]

Two formal differences capture the distinction between the conventional model and this generalization of it. First, individual type is now defined by a duple $\{\theta_i, w_i\}$, where the preference parameter θ_i deflates the quantity of labor supplied. In the conventional model, $\theta_i = 1$ for all $i \in \{1, \dots, I\}$. Here, "experienced labor" l_i/θ_i denotes the relative preference-adjusted amount of labor effort by individual i. Second, the utility function in (1) is type-specific, because utility as a function of consumption and labor effort now depends on the individual's preference parameter θ_i. In the conventional setup, heterogeneity enters only

[7] As expression (1) makes clear, one could, as in the standard approach, include both components of the heterogeneity in unified type into the individual's budget constraint. Instead, our approach treats the two components of unified type differently, with preference heterogeneity affecting the utility function directly. This facilitates an intuitive understanding of why that component may not justify redistribution, in contrast to differences in ability that affect individuals' budget constraints.

through the budget constraint and utility functions are homogeneous.

This formalization of preferences allows us to collapse the two dimensions of individual heterogeneity into a single "unified type" $n_i = \theta_i w_i$, such that any joint distribution of preferences and ability $\Phi(\theta, w)$ generates a univariate distribution of unified type $F(n)$ with support over the range $[n_0, \infty)$. This unified type serves as a one-dimensional measure of heterogeneity that determines i's observable behavior. It allows us to analyze optimal policy exactly as in the conventional approach, where the required technical conditions (especially the single-crossing property) are now assumed to hold with respect to unified type rather than ability.[8] This compression of multiple dimensions of heterogeneity into one is a particularly simple example of the technique used in Choné and Laroque (2010).

This unification of types also provides a straightforward way to formalize the social planner's response to the normative complications raised by preference heterogeneity.[9] We now turn to the planner's problem.

1.2 Social Planner

The social planner maximizes social welfare by specifying taxes as a differentiable function of income, denoted $T(y) : \mathbb{R}_{\geq 0} \to \mathbb{R}$, where income is a function of unified type $n > 0$. Social welfare W is the integral of the product of a function $G(n)$ and scalar weights α_n. The function $G(n)$ represents the social value of the utility of an individual of type n, and we define $g(n) = dG(n)/dc(n)$ as the marginal social value of consumption for that individual.[10] We write $G(n)$ and $g(n)$ as functions of n to maximize the generality of the model; this need not imply that they depend on agents' types directly. We assume $g(n)$ decreases in n, as is standard, providing a motive for redistribution from high types to low. To manage

[8] A related observation is that the set of Pareto-efficient allocations is identical under the conventional model and our model when the distribution of abilities in the conventional model is the same as that of the unified types in our model. When choosing the optimal allocation within the Pareto-efficient set, however, any normative distinction between ability and unified type becomes important.

[9] Though preferences and ability are observationally equivalent and unobservable, we believe there is a meaningful *positive*, as well as normative, distinction between them. Unobservability of preferences is public, not private; that is, it would not be incentive compatible to reveal one's true preferences to the planner. That preferences are privately understood as different from ability, and that aggregate information on preference heterogeneity might be useful, is suggested by evidence from the opinion surveys discussed in Section 3.2.

[10] Familiar specifications of these functions in conventional models are, for example, $G(n) = U(n)$ and $g(n) = U_c(n)$ for a utilitarian planner or $G(n) = G(U(n))$ and $g(n) = G'(U(n))U_c(n)$ with $G''(U(n)) < 0$ for a more redistributive planner.

the addition of preference heterogeneity, the $G(n)$ functions are weighted by type-specific adjustment factors α_n. Below, we describe how the values of α_n are chosen.

Maximization is subject to standard feasibility and incentive compatibility constraints. Formally, the planner's problem is:

Problem 1 *Planner's Problem*

$$\max_{T(y)} W = \int_{n=0}^{\infty} \alpha_n G(n) f(n) \, dn, \tag{2}$$

where $\alpha_n \geq 0$ *for all* $i \in \{1, \ldots, I\}$ *are weights on the social welfare values of individual utilities, subject to feasibility:*

$$\int_{n=0}^{\infty} T(y(n)) f(n) \, dn = 0 \tag{3}$$

and incentive compatibility:

$$y(n) = \arg\max_{y} \left[U\left(y - T(y), \frac{y}{n} \right) \right] \; \forall n. \tag{4}$$

A tax function is *feasible and incentive compatible* if it satisfies conditions (3) and (4).

This social planner's problem is formally similar to that in a conventional analysis with a generalized utilitarian planner, but the weights $\{\alpha_n\}$ play a novel role in this setting. For clarity, we will call each α_n the "scaling factor" for unified type n. Given a vector of such scaling factors, or simply a "scaling", the planner values marginal consumption for type m relative to that for another type n as follows:

$$\frac{\partial W / \partial c_m}{\partial W / \partial c_n} = \frac{\alpha_n g(n)}{\alpha_m g(m)}.$$

Naturally, these relative valuations have substantial effects on optimal policy.

As the scaling $\{\alpha_n\}$ is empirically unidentified, its selection is a normative decision. In a conventional model, type is fully determined by ability, and the choice of α_i is relatively straightforward. In Saez (2001), for example, the planner simply sets $\alpha_n = 1$ for all n. In this model, however, the set of individuals of any given unified type n may vary in the two underlying sources of heterogeneity, namely preferences θ and ability w. If these sources of heterogeneity have different normative significance for the planner, the planner's values for the scaling $\{\alpha_n\}$ will be affected. Without loss of generality, we will restrict our consideration to scalings such that $E[\alpha_n] = 1$, since multiplying all values of $g(n)$ by a common factor does not affect the optimal policy. Pinning down the appropriate values for this scaling has been an obstacle to definitive results on preference heterogeneity and optimal redistribution. Fortunately, work by Fleurbaey and Maniquet (2006) has found a way around that obstacle.

We follow Fleurbaey and Maniquet (2006) in imposing a normative requirement that we label *preference neutrality*. Defining preference neutrality requires a preliminary step. Consider the *laissez-faire* tax regime, in which individuals retain all the earnings they generate.[11] In that setting, individual i maximizes utility in (1) subject to the *laissez-faire* budget constraint $c_i = y_i$. Let $y_i^{LF} \geq 0$ denote individual i's choice in this setting. Then, $y_i^{LF} = y^{LF}(\theta_i w_i)$ is the implicit function defined by the agent's first-order condition:

$$U_c\left(y^{LF}, \frac{y^{LF}}{\theta_i w_i}\right) + \frac{U_l\left(y^{LF}, y^{LF}/(\theta_i w_i)\right)}{\theta_i w_i} = 0,$$

where U_c and U_l denote the partial derivatives of utility with respect to consumption and labor effort. Our goal is to select a scaling such that this tax regime is optimal when ability is uniform, that is, when all income inequalities arise from differences in preferences. Let $g^{LF}(n)$ denote the marginal social value of consumption for n under this tax regime.[12]

Using this information, we define preference neutrality as follows.

Definition 1 *The scaling $\{\alpha_m\}$ satisfies "preference neutrality" if and only if*

$$\frac{\alpha_m E[g^{LF}(\theta_i \bar{w})|\theta_i w_i = m]}{\alpha_n E[g^{LF}(\theta_j \bar{w})|\theta_j w_j = n]} = 1, \ \forall m, n, \tag{5}$$

where $\bar{w} = E[w_i]$, the mean wage in the economy.[13]

[11] The laissez-faire is a hypothetical regime, and thus unobservable, but that does not prevent us from deriving the analytical or empirical results in this paper.

[12] The superscript LF is necessary, as $g(\cdot)$ may be endogenous to the tax regime in general. That is, unlike the $g(\cdot)$ terms in most cases, $g^{LF}(\cdot)$ is independent of the α_i values because the tax system in the laissez-faire is fixed (i.e., there is no taxation).

[13] As demonstrated in the appendix, for a common class of examples the choice $\bar{w}$ is irrelevant.

In words, preference neutrality adjusts the marginal social weights α_n so that inequalities that would arise due to preferences if all agents had a common ability do not merit redistribution.[14] In the conventional Mirrlees case of homogeneous preferences, $\theta_i = 1$ for all i. In that case, (5) becomes $\frac{\alpha_m g^{LF}(\bar{w})}{\alpha_n g^{LF}(\bar{w})} = 1$, implying $\alpha_m = \alpha_n$, and the standard Mirrlees results hold. On the other hand, if ability is homogenous, then (5) becomes $\frac{\alpha_m g^{LF}(m)}{\alpha_n g^{LF}(n)} = 1$, implying the *laissez-faire* allocation is optimal. Since $g^{LF}(n)$ decreases with n, $g^{LF}(\theta\bar{w})$ is decreasing in θ, and therefore if the conditional distribution $F(\theta_i|\theta_i w_i = m)$ first-order stochastically dominates $F(\theta_i|\theta_i w_i = n)$ whenever $m > n$ (intuitively, if the distribution of preferences is "rising" with incomes) then α_n is rising with n.

With the objective in (2), condition (5) means that under preference neutrality, preference heterogeneity alone does not justify redistribution. That idea has been understood for some time, and is clearly stated in (for example) Sandmo (1993), but reliable implications of it for optimal policy have been elusive. We now turn to the task of deriving the implications of allowing this type of preference heterogeneity into a standard optimal tax model.

2 A sufficient condition for preference heterogeneity to reduce redistribution

In this section, we compare the tax policies that solve the planner's problem of Section 1 under two possible assumptions about the sources of heterogeneity. The benchmark assumption is that the distribution of unified types $F(n)$ is produced by a degenerate joint distribution of preferences and ability in which $\theta_i = 1$ and $n_i = w_i$, for all $i \in \{1, \ldots, I\}$. In contrast, the alternative assumption studied in this paper allows some heterogeneity in income to be attributed to heterogeneous preferences, so that $\theta_i \neq \theta_j$ for some types i, j.

We compare policies along two dimensions–progressivity and redistribution, which we formally define as follows.

[14] The preference-neutral marginal social weights depend on the joint distribution of abilities and preferences. This endogeneity is due to our assumption that θ_i is unobservable, so that the α_i terms involve a conditional expectation over the unified type distribution. If preferences were observable, these social marginal weights would depend only on each individual's allocation, just as in the standard model.

Definition 2 *Consider two feasible and incentive compatible, Pareto efficient income tax functions* $T(y): \mathbb{R}_{\geq 0} \to \mathbb{R}$ *and* $\hat{T}(y): \mathbb{R}_{\geq 0} \to \mathbb{R}$*, resulting in endogenous income distributions* $y(n)$ *and* $\hat{y}(n)$*.*

- $T(y)$ *is "less progressive" than* $\hat{T}(y)$ *if marginal tax rates under* $T(y)$ *are less than those under* $\hat{T}(y)$ *for all agents, that is if* $T'(y(n)) < \hat{T}'(\hat{y}(n))$ *for all* n*;*
- $T(y)$ *is "less redistributive" than* $\hat{T}(y)$ *if the net tax on the lowest type agent under* $T(y)$ *is greater than that under* $\hat{T}(y)$*, that is if* $T(y(n_0)) > \hat{T}(\hat{y}(n_0))$*.*

Given these definitions, we can show a useful and intuitive lemma, proven in the Appendix.

Lemma 1 *If the tax function* $T(y)$ *is less progressive than the tax function* $\hat{T}(y)$*, then* $T(y)$ *is less redistributive than* $\hat{T}(y)$*.*

To analyze the implications of preference heterogeneity for optimal tax policy, we use Diamond's (1998) well-known expression for optimal marginal tax rates when agent utility is quasilinear in consumption: that is when

$$U(c(n), y(n)/n) = c(n) + v(y(n)/n), \tag{6}$$

with $v'(l) < 0$, $v''(l) < 0$. This case is particularly tractable, and it allows us to prove a sharp result that we formalize in the following proposition.[15] We have been unable to extend this result to the case of nonzero income effects of tax changes—that is, when individuals' labor supply decisions are affected by changes to marginal tax rates on incomes lower than their own.[16] We therefore rely on numerical simulations to confirm that the proposition's findings extend to that more general case.

[15] Technically, the absence of income effects means a change in the lump sum grant does not induce a change in labor supply, facilitating the proof. As Diamond notes, "In the presence of distorting taxes, income effects imply that lump-sum taxes have efficiency effects since they change distorted labor supply decisions."

[16] Analysis toward sufficient conditions for the case of income effects is available from the authors.

Incorporating our preference neutral scaling into Diamond's (1998) condition for the optimal marginal income tax rate yields

$$\frac{T'(y(n))}{1-T'(y(n))} = \frac{1+\varepsilon}{\varepsilon n f(n)} \int_n^\infty \left(1 - \frac{\alpha(m)g(m)}{\lambda}\right) dF(m), \tag{7}$$

where ε denotes the elasticity of taxable income, and λ denotes the shadow value of public funds. In the conventional approach, n represents ability; here we interpret n as unified type and use the preference neutral scaling factors $\alpha(n)$ to adjust for preference heterogeneity. Note that we could incorporate the α_n weights into the $G(n)$ functions. By separating them out, however, we can use the α_n terms to absorb any effects of preference heterogeneity on the planner's objective. In particular, in this section we assume the $g(n)$ values are independent of the α_n values.[17] This independence is useful in proving the following proposition.

Proposition 1 *Let $T(y)$ solve the planner's problem in (2) through (4) in the presence of preference heterogeneity, and let $\hat{T}(y)$ solve this problem with $\theta_i = 1$, $\forall i$. If $F(\theta|y_i)$ stochastically dominates $F(\theta|y_j)$ whenever $y_i > y_j$, and if individual utility takes the quasilinear form in (6), then $T(y)$ is less progressive and less redistributive than $\hat{T}(y)$.*

Proof.

The planner values consumption for unified type n at $\alpha(n)g(n)$. At the optimum, $\lambda = \int_0^\infty \alpha(m)g(m)dF(m)$, since perturbing the demogrant must have no first-order effect on social welfare. If the true distribution of preferences is heterogeneous such that $F(\theta|y_i)$ first-order stochastically dominates $F(\theta|y_j)$ whenever $y_i > y_j$, then since y is increasing with unified type n, α_n is increasing in n. Under the Mirrlees benchmark, tastes are mistakenly assumed to be constant, with scaling factors equal to one. Let $\tilde{\alpha}(n) = \alpha_n - 1$ denote the difference between the true preference neutral scaling α_n and the Mirrlees benchmark; note $\tilde{\alpha}(n)$ is increasing in n and $E[\tilde{\alpha}(n)] = 0$. Then we can let $\rho \in [0,1]$ parameterize the continuous

[17] The simplest form for $G(n)$ yielding such independence is to set $G(n)$ equal to scalars that depend on type n only, a frequently-used approach (see Salanié 2011, for a textbook treatment). Given quasilinear utility, in this case the terms $g(n)$ also depend only on type n, and $g'(n) < 0$ is assumed. As long as these weights decline so that incentive constraints bind in the typical way, this setting yields results comparable to models with strictly concave social marginal utility of consumption, such as the standard utilitarian setup. A more nuanced form for $G(n)$ also yielding such independence is as follows. Define $h(n)$ as the difference between the utility levels of individual n under the optimal policy (i.e., given values for the α_n that adjust for preference heterogeneity) and under the optimal policy given the conventional values $\alpha_n = 1$ for all n. Then define $G(n) = G(U(n) - h(n))$, which we assume to be concave. By construction, $U(n) - h(n)$ obtains the same value for different values of α_n. Given quasilinear utility, $g(n) = G'(U(n) - h(n))$ is therefore independent of α_n, and $g'(n) < 0$.

transformation from the Mirrlees benchmark to the true optimum:

$$\frac{T'(y(n))}{1-T'(y(n))} = \frac{1+\varepsilon}{\varepsilon n f(n)} \left(1 - F(n) - \frac{\int_n^\infty (1+\rho\tilde{\alpha}(m))g(m)dF(m)}{\int_0^\infty (1+\rho\tilde{\alpha}(m))g(m)dF(m)}\right), \tag{8}$$

We wish to sign the derivative of this expression with respect to ρ—this will indicate the directional impact on marginal tax rates of perturbing the Mirrlees benchmark toward the true optimum under preference heterogeneity. By construction, $g(n)$ does not depend on preferences, so $g(n)$ does not vary with ρ. Therefore the derivative of (8) with respect to ρ, evaluated at $\rho = 0$, is

$$\left(\frac{1+\varepsilon}{\varepsilon n f(n)}\right) \frac{\int_n^\infty g(m)dF(m) \cdot \int_0^\infty \tilde{\alpha}(m)g(m)dF(m) - \int_0^\infty g(m)dF(m) \cdot \int_n^\infty \tilde{\alpha}(m)g(m)dF(m)}{\left(\int_0^\infty g(m)dF(m)\right)^2}.$$

The denominator is positive, so the sign is determined by the numerator, which can be manipulated as follows:

$$\begin{aligned}
&\int_n^\infty g(m)dF(m) \cdot \int_0^\infty \tilde{\alpha}(q)g(q)dF(q) - \int_0^\infty g(r)dF(r) \cdot \int_n^\infty \tilde{\alpha}(s)g(s)dF(s) = \\
&\int_0^\infty \int_n^\infty g(m)\tilde{\alpha}(q)g(q)dF(m)dF(q) - \int_0^\infty \int_n^\infty g(r)\tilde{\alpha}(s)g(s)dF(s)dF(r) = \\
&\int_0^\infty \int_n^\infty \left(g(m)\tilde{\alpha}(q)g(q) - g(q)\tilde{\alpha}(m)g(m)\right) dF(m)dF(q) = \\
&\int_0^\infty \int_n^\infty g(m)g(q)\left(\tilde{\alpha}(q) - \tilde{\alpha}(m)\right) dF(m)dF(q) = \\
&\int_0^n \int_n^\infty g(m)g(q)\left(\tilde{\alpha}(q) - \tilde{\alpha}(m)\right) dF(m)dF(q) + \int_n^\infty \int_n^\infty g(m)g(q)\left(\tilde{\alpha}(q) - \tilde{\alpha}(m)\right) dF(m)dF(q).
\end{aligned}$$

The right term integrates to zero. If $\tilde{\alpha}(n)$ is increasing (decreasing) in n, the left term is negative (positive). Therefore the derivative of (8) with respect to ρ is negative for all n—that is, perturbing the Mirrlees benchmark toward the true optimum lowers all marginal tax rates. The same argument holds at all $\rho \in (0,1)$, which, together with Lemma 1, implies Proposition 1.

■

In words, Proposition 1 says that the true optimal policy $T(y)$ features smaller marginal tax rates and less redistribution than the Mirrlees benchmark policy $\hat{T}(y)$ whenever consumption preferences rise, in the sense of first-order stochastic dominance, with income. It is important to note that this sufficient condition depends on the relationship between preferences and income—for which it seems highly plausible—not preferences and ability.

The intuition for Proposition 1 is as follows. If income variation is due to variation in both preferences for consumption relative to leisure (which does not merit redistribution) and ability (which does), then the optimal extent of redistribution depends on how preferences vary with income. If preferences and income move together, then attributing all of a high earner's income to ability would exaggerate his or her ability level and yield too great an extent of optimal redistribution. If, on the other hand, preferences fall with income, then high earners must have greater abilities than a model without preference heterogeneity would infer. In that case, optimal redistribution rises relative to a conventional analysis. Proposition 1 clarifies a condition on the relationship between preferences and income that guarantees the former case holds, i.e., that preference heterogeneity lowers optimal redistribution.

Proposition 1 establishes a new result that helps to resolve the theoretical ambiguity over the effects of preference heterogeneity on optimal tax policy. To build on this result and obtain quantitative implications, we now turn to a more fully specified model.

3 A simple statistic for the quantitative effects of preference heterogeneity on redistribution

In this section, we choose specific functional forms–for the social welfare function $G(n)$, individual utility, and the distributions of ability and preferences–enabling us to derive a second novel analytical result: a simple statistic for the effect of heterogeneity in preferences on optimal marginal tax rates. We assume the planner is a "pure Utilitarian":

$$G(n) = U(n). \tag{9}$$

so that $g(n) = \frac{\partial U_i(c_i,l_i)}{\partial c_i}$, that is, the social marginal value of consumption is the unscaled individual marginal utility of consumption. We assume that the utility function in expression (1) takes the following familiar form:

$$U_i(c_i, l_i) = \frac{c_i^{1-\gamma} - 1}{1-\gamma} - \frac{1}{\sigma}\left(\frac{l_i}{\theta_i}\right)^{\sigma}, \tag{10}$$

where $\gamma > 0$, the coefficient of relative risk aversion, determines the concavity of utility from consumption (for which $\gamma = 1$ is taken to represent log utility from consumption), $\sigma \geq 1$ affects the elasticity of labor supply, and $l_i = y_i/w_i$. We will further suppose that $\ln \theta_i$ and $\ln w_i$ are jointly normally distributed, so that y^{LF} has a lognormal distribution. Though evidence (see Saez 2001) suggests that the upper tail of the income distribution is better described as a Pareto distribution, lognormality has long been used in the optimal tax literature to describe most of the income distribution (see Tuomala 1990). Moreover, assuming preferences are bounded above, the optimal top marginal tax rate in this generalized setting will be the same as that obtained in the conventional analysis with no preference heterogeneity. The effects of preference heterogeneity on tax rates is thus more relevant for the mass of the earnings distribution below the top tail, a mass well-described as lognormal.

Given these conditions, we can prove the following result relating the joint distribution of preferences and income to the optimal tax structure.

Proposition 2 *Let $T(y)$ solve the planner's problem in (2) through (4). If the planner is utilitarian and individual utility takes the form in (10), and $\ln \theta_i$ and $\ln w_i$ are jointly normally distributed, then $T(y_i)$ depends on the distribution of preferences $\{\theta_i\}_i$ only through the statistic β :*

$$\beta = \frac{\text{cov}(\ln \theta_i, \ln n_i)}{\text{var}(\ln n_i)}. \tag{11}$$

Proof.

Under the utility function in (10), $y^{LF}(\theta_i w_i) = (\theta_i w_i)^{\frac{\sigma}{\sigma+\gamma-1}}$, and $g^{LF}(n) = \left(y^{LF}(n)\right)^{-\gamma}$, so, from (5), the preference neutral scaling and the scaling $\{\alpha\}$ satisfies

$$\frac{\alpha_m}{\alpha_n} = \frac{E\left[g^{LF}(\theta_i \bar{w})|n_i\right]}{E\left[g^{LF}(\theta_i \bar{w})|m_i\right]} = \frac{E\left[(\theta_i \bar{w})^{\frac{-\gamma\sigma}{\sigma+\gamma-1}}|n_i\right]}{E\left[(\theta_i \bar{w})^{\frac{-\gamma\sigma}{\sigma+\gamma-1}}|m_i\right]} = \frac{E\left[\theta_i^{\frac{-\gamma\sigma}{\sigma+\gamma-1}}|n_i\right]}{E\left[\theta_i^{\frac{-\gamma\sigma}{\sigma+\gamma-1}}|m_i\right]}. \tag{12}$$

Note that under this utility function, the preference neutral scaling does not depend on the mean wage $\bar{w}$.[18]

[18] The irrelevance of $\bar{w}$ holds under utilitarianism whenever individual utility exhibits constant relative risk aversion and labor supply elasticity is uniform in the *laissez-faire* state. This is proved in Lemma 2 in the Appendix.

Because the conditional distribution of $\ln\theta_i$ is normal, the numerator can be written

$$\begin{aligned} E\left[\theta_i^{\frac{-\sigma\gamma}{\sigma+\gamma-1}}|n_i\right] &= E\left[\exp\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\ln\theta_i\right)|n_i\right] \\ &= \exp\left(E\left[\frac{-\sigma\gamma}{\sigma+\gamma-1}\ln\theta_i|n_i\right]\right)\cdot\exp\left(\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\right)^2\cdot\frac{\operatorname{var}(\ln\theta_i|n_i)}{2}\right) \qquad (13) \end{aligned}$$

By construction, $\ln n_i$ and $\ln\theta_i$ are jointly normally distributed, and therefore the conditional variance $\operatorname{var}(\ln\theta_i|\ln n_i)$ is independent of $\ln n_i$. Therefore the right exponentiated factor in (13) is constant with respect to n_i—call its value K.

When X and Y are jointly normal, $E[X|Y] = E[X] + (cov(X,Y)/var(Y))(Y - E[Y])$, so we rewrite the expectation in (13) as

$$\begin{aligned} \frac{-\sigma\gamma}{\sigma+\gamma-1}E[\ln\theta_i|\ln n_i] &= \frac{-\sigma\gamma}{\sigma+\gamma-1}\left(E[\ln\theta] + \frac{\operatorname{cov}(\ln\theta_i,\ln n_i)}{\operatorname{var}(\ln n_i)}\cdot(\ln n_i - E[\ln n])\right) \\ &= \frac{-\sigma\gamma}{\sigma+\gamma-1}\left(E[\ln\theta] + \beta\ln n_i - \beta E[\ln n]\right). \qquad (14) \end{aligned}$$

Substituting (14) into (13) and then into (12), we have:

$$\begin{aligned} \frac{\alpha_m}{\alpha_n} &= \frac{\exp\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\left(E[\ln\theta] + \beta\ln n - \beta E[\ln n]\right)\right)\cdot K}{\exp\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\left(E[\ln\theta] + \beta\ln m - \beta E[\ln n]\right)\right)\cdot K} \\ &= \frac{\exp\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\beta\ln n\right)}{\exp\left(\frac{-\sigma\gamma}{\sigma+\gamma-1}\beta\ln m\right)} = \left(\frac{m}{n}\right)^{\beta\left(\frac{\sigma\gamma}{\sigma+\gamma-1}\right)}. \end{aligned}$$

In particular, the scaling $\alpha_n = n^{\beta\left(\frac{\sigma\gamma}{\sigma+\gamma-1}\right)}$ is preference neutral; plugging it into the well-known condition derived in Saez (2001) yields the optimal nonlinear income tax.[19] Thus the population statistic β is sufficient to incorporate fully the effect of preference heterogeneity in this case.

■

This simple and intuitive[20] statistic proves to be a convenient tool with which to examine the quantitative, not just qualitative, implications of preference heterogeneity for policy. To illustrate this, note that $\beta = 0$ corresponds to the Mirrlees case of homogeneous preferences and $\beta = 1$ corresponds to the opposite extreme of homogeneous ability. In the latter case, $\alpha_n = n^{\left(\frac{\sigma\gamma}{\sigma+\gamma-1}\right)}$ for all n, and the optimal extent of redistribution is zero. If $\beta < 0$, optimal

[19] Of course, any other scaling that yields the same ratios α_m/α_n for all m and n would generate the same optimal income tax, as this would amount to multiplying all agents' utility functions by the same constant.

[20] Note that β can be interpreted as the coefficient from a regression of $\ln\theta$ on $\ln n$.

redistribution is greater than in the conventional analysis.[21]

We can also write β as the product of the correlation between $\ln\theta$ and $\ln n$ and the ratio of their standard deviations. Thus, a particularly simple case obtains when preferences and ability are independent. Then $\text{var}(\ln n) = \text{var}(\ln\theta) + \text{var}(\ln w)$, and $\beta = \text{var}(\ln\theta)/\text{var}(\ln n)$, while $1-\beta = \text{var}(\ln w)/\text{var}(\ln n)$. That is, when preferences and ability are independent, β represents the "share" of of the variation in income driven by preference heterogeneity.

3.1 Numerical simulations of optimal policy

We now use calibrated numerical simulations and the statistic β to characterize the potential quantitative effects of preference heterogeneity on optimal policy.

Our calibration strategy is to match the income distribution chosen by individuals in this Section's model, taking U.S. tax policy as given, to the empirical income distribution in the United States, and thus infer a distribution of unified types $F(n)$. We assume a pure Utiltarian social planner. We consider a range of utility function parameter values for expression (10) that spans most mainstream estimates: $\gamma \in \{0.5,\ 1,\ 2\}$ and $\sigma \in \{2,\ 3,\ 6\}$. We use data from the Tax Policy Center reporting effective marginal and average tax rates at the following income percentiles: 20, 40, 60, 80, 90, 95, 99, and 99.9. Using the utility function (10), and given values for γ and σ, we can back out the implied unified type n_i at each quantile. We assume that the unified types n are drawn from a lognormal distribution with parameters μ and ξ. Then, we search for the values of μ and ξ that minimize the sum of squared differences between the resulting simulated income quantile thresholds and the empirical ones. For example, using the parameter values $\gamma = 1$ and $\sigma = 3$, the resulting parameter estimates are $\mu = 1.6673$ and $\xi = 1.0025$. Our conceptual results are not sensitive to these values, but having a realistic calibration makes the magnitudes of our results easier to interpret.

We then use this calibrated model to calculate optimal policy under a range of values for β as defined in (11). We choose values for the welfare weights α_i to satisfy Preference

[21] This discussion of Proposition 2 shows that it is closely related to Proposition 1 from the previous section. If preferences rise, in the sense of first-order stochastic dominance, with income, then the covariance between preferences and unified type (equivalently, income) is positive, and therefore $\beta > 0$.

Neutrality and solve for optimal policy according to the planner's problem in (2) through (4). Figure 1 plots average tax rates against income for four values of β and a baseline set of utility parameters: $\gamma = 1$ and $\sigma = 3$. It shows how optimal redistribution varies with the role attributed to preference heterogeneity.

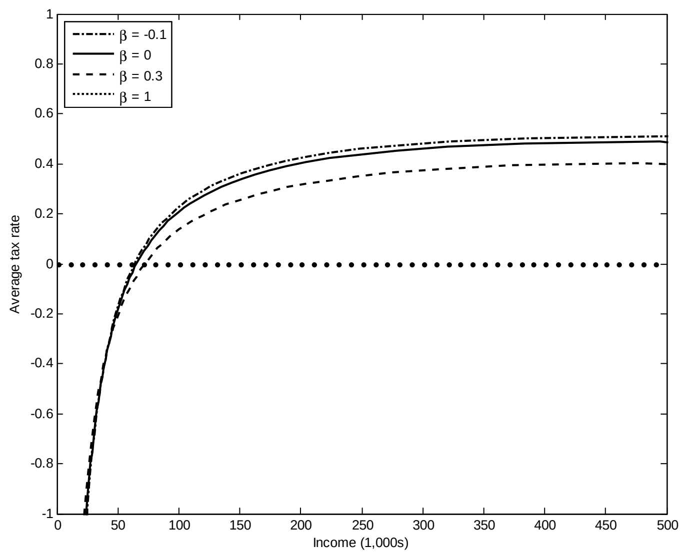

Figure 1: Optimal average tax rate schedules for four values of β

In Figure 1, the Mirrleesian benchmark is the $\beta = 0$ scenario, where all heterogeneity is ascribed to ability. The opposite assumption, where all heterogeneity is ascribed to preferences, is $\beta = 1$. An intermediate value of $\beta = 0.3$ is used to provide a sense for the effects of a more moderate adjustment to the standard model. Finally, we show the results for $\beta = -0.1$, a negative value. A negative β indicates that high incomes are associated with low preferences for consumption (and thus very high abilities).[22]

The results for positive values of β suggest that preference heterogeneity has the potential to substantially reduce the optimal extent of redistribution across a given income

[22]It is theoretically possible for β to be greater than one, as well, which would indicate that high incomes are associated with very high preferences for consumption (and thus very low abilities). If $\beta > 1$, optimal policy is even less redistributive than the $\beta = 1$ case shown in Figure 1. This case is also considered in Chonè and Laroque (2010) in their analysis of when optimal marginal tax rates are negative.

distribution. For example, while the conventional case of $\beta = 0$ recommends a tax rate of 49 percent at \$500,000 of annual earnings, the case with $\beta = 0.3$ recommends a rate of 40 percent. For reference, the empirical average tax rate at \$500,000 in 2012 was approximately 27 percent. The results for $\beta = -0.1$ confirm the role of the sufficient condition in Proposition 1, as in that case preference heterogeneity increases the optimal extent of redistribution and preferences for consumption do not rise with income.

Figure 2 shows the average tax rate schedules for the values of β across nine combinations of γ and σ, where the center graph is identical to Figure 1 and is provided for comparison purposes.

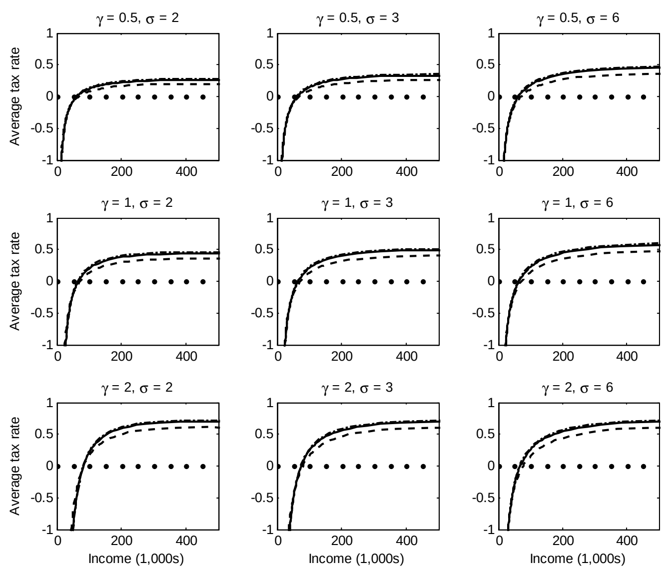

Figure 2: Optimal average tax rate schedules for a range of parameter values. In each case, the solid line shows $\beta = 0$, the conventional benchmark. The dotted line shows $\beta = 1$, the dashed line shows $\beta = 0.3$, and the dash-dot line shows $\beta = -0.1$.

The results in Figure 2 demonstrate that the effect of preference heterogeneity is not sensitive to the particular values selected for γ and σ. They also suggest that these effects, and therefore the conclusions of Proposition 1, are not sensitive to the inclusion of income effects.

Though β is not conventionally observable, these simulations show its value as a straightforward way in which to modify a numerical version of the model to determine the potential quantitative implications of preference heterogeneity. Moreover, β may be a more plausible empirical target than has previously been identified, especially if unconventional sources of evidence are brought to bear, as we now show.

3.2 Suggestive empirical patterns

To demonstrate the empirical potential of our results, and to reinforce the usefulness of the population statistic β, we now provide some suggestive evidence that preference heterogeneity may be related to real-world policy across countries and U.S. states in the ways that our previous results suggest. We emphasize that these results are admittedly far from conclusive and are vulnerable to a variety of criticisms. Our hope is that they stimulate further data gathering and empirical work that can more reliably test for the implications of the theory in existing policy.

First, we consider cross-sectional[23] international data. The World Values Survey asked the following question of respondents between 2005 and 2007: *"Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? ...It is important to this person to be rich; to have a lot of money and expensive things."* We will use the answers to the question to measure preferences $\{\theta_i\}$. This question is well-posed for our purposes, as it attempts to have the respondent reflect on his or her underlying preferences rather than how he or she feels in the status quo, i.e., "on the margin." Moreover, the World Values Survey's international coverage is unmatched. The World Values Survey also asks respondents to report their place in the income distribution (it asks which of ten "steps" the respondent's household income falls into). Since income increases monotonically

[23] 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.

with unified type, we use these reported values as a measure of unified type $\{n_i\}$.[24] It is possible to calculate the covariance of these data within each country, giving us all of the components required to calculate β. We relate these values of β to a standard measure of redistribution, the level of social (transfer) expenditures as a share of GDP, as reported by the OECD. Such expenditures include benefits for the poor, disabled, and elderly, as well as health-related and other transfer programs. We are able to calculate β for 13 countries with PPP-adjusted GDP per capita greater than $25,000 in 2005 U.S. dollars, a simple threshold that helps to control for the wide range of institutional variables that likely affect the scale of redistribution. Figure 3 shows the results visually.

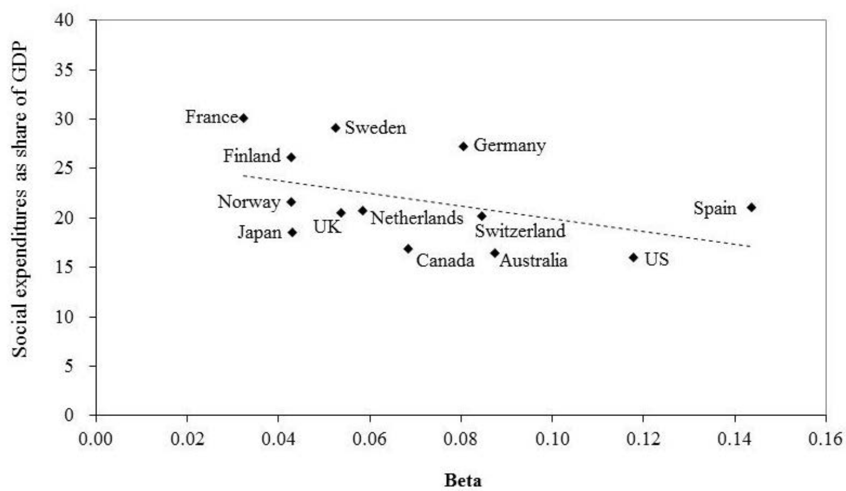

Figure 3: Redistribution and β in 13 rich countries

[24] The distribution of responses is far from uniform across deciles in most countries, which could reflect either the pattern of sampling or surveyor and respondent behavior toward such a question. Though their exact meaning is ambiguous, these data are arguably superior even to perfect information on individuals' incomes (which, regardless, we lack) because we are using them in concert with individuals' stated attitudes toward the importance of being "rich." Income data that reflect a respondent's perceived place in the income distribution may provide a more appropriate match to these preference data. Related, we calculate β in this analysis using the levels of preferences and income reported, not the logs of those levels. The reason is that income is reported on a linear scale from 1 to 10, so that individuals are likely interpreting the scale roughly as deciles. If income is distributed roughly log-normally, taking logs of these levels would be redundant. For consistency, we assume that the preference scale represents a similar implicit transformation. Nevertheless, if we calculate β using the logs of these scales instead, the sign and significance of the coefficient on β are virtually unchanged from the analysis shown.

A weak but noticeable negative relationship between redistribution and β is apparent in Figure 3, consistent with the theory developed above. That is, countries in which preference heterogeneity plays a larger role in explaining income variation appear to have less redistributive policies. Note that this relationship may reflect interdependence rather than unidirectional causality.[25] The point estimate of the coefficient on β is -64.1, and it is significant at the 15 percent level (with a standard error of 39.9); it is also the slope of the dashed best-fit line shown in the figure. Though this evidence is far from definitive, the relationship shown here is robust to controlling for some obvious alternative explanations. If we control for the log of GDP per capita and the extent of inequality as measured by the pre-tax Gini coefficient, the size and significance of the coefficient on β slightly increase, namely to -66.0 significant at the the 12.5 percent level. The results weaken somewhat, to a coefficient of -49.4 significant at only the 35 percent level) if we include three OECD countries at an earlier stage of development than those shown in Figure 3–Chile, Korea, and Poland–the first two of which have social expenditures of 10 and 7 percent of GDP but low values of β. Of course, any results with such a limited sample are merely suggestive of a relationship that, given the potential feasibility of measuring the statistic β, may reward greater study.

Next, we turn to data for states within the United States. For preferences, as with the international data, we use responses to a question on the importance of material possessions. The General Social Survey, administered in 1993, asked *"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?"* Unfortunately, restrictions on the data prevent us from calculating values for β because we cannot link respondents preferences and income.[26] As an alternative, we compare redistribution to the ratio of the standard deviation of preferences to the standard deviation of log income in a state.[27] Recall that β equals this ratio multiplied

[25] That is, it may be that residents of countries with more redistributive policies tend to evolve toward having more similar preferences. Related, it may be of interest to note that the pattern in Figure 3 is consistent with Alesina and Angeletos (2005). In redistributive countries, small β implies that ability–"luck" in their model–plays a large role in determining income relative to preferences–"effort" in their model.

[26] 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.

[27] We estimate the variance of log income using state-level annual wages at the 10th, 25th, 50th, 75th, and 90th percentiles, obtained from the Bureau of Labor Statistics (http://www.bls.gov/oes/current/oessrcst.htm). We select the Pareto log-normal distribution that

by the correlation between preferences and income. To the extent that this correlation varies across states independently of the ratio we are able to calculate, our approach will give a reliable indication of the relationship between redistribution and β. For redistribution we follow 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. Figure 4 shows the results (confidentiality prevents us from identifying the states).

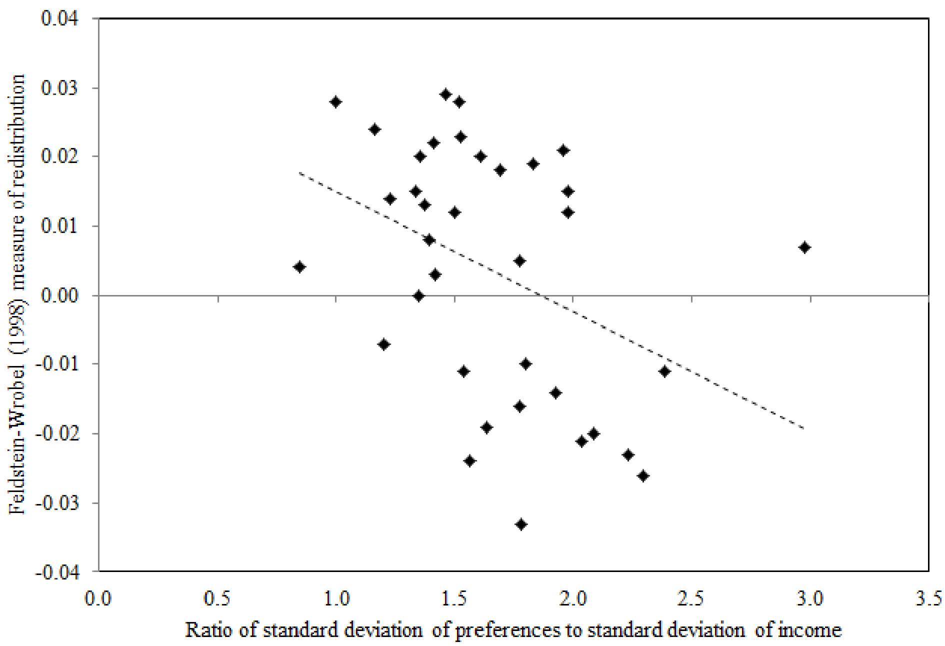

Figure 4: Redistribution and a proxy for β in 36 U.S. states

As Figure 4 shows, the same suggestive pattern as was apparent across countries appears to hold across U.S. states. The coefficient on the ratio of the standard deviation of preferences to the standard deviation of income is -0.017, which is significant at the 2 percent level (it is the slope of the dashed best-fit line shown in the figure). If we control for the log of average personal income in each state and the share of state income claimed by the top decile of earners (a proxy for inequality), the magnitude of the estimated coefficient and its significance both slightly increase.

best matches these observed data (in terms of minimum squared errors at the observed values of the distribution function), then calculate the variance of log incomes for each calibrated distribution.

We reiterate that these empirical patterns are meant to suggest, not prove, that the relationship between preference heterogeneity and redistribution identified in the theory and captured (under specific functional form assumptions) by the statistic β may be consistent with variation in existing policy. At the least, these analyses demonstrate that further empirical study of preference heterogeneity's relationship to prevailing policies may be both possible and rewarding.

4 Conclusion

By adopting and adding to recent innovations in optimal tax analysis, we have shown that long-standing intuitions about preference heterogeneity reducing optimal redistribution are incomplete but correct given a plausible condition on how preferences relate to income. We have also shown that a simple statistic for this effect exists, assuming familiar functional forms for utility and the distributions of ability and preferences, and we have shown how that statistic provides a straightforward and potentially empirically-relevant way to gauge the quantitative implications of preference heterogeneity for redistribution.

Though these results modify those of the conventional Mirrleesian approach, we see them as essential steps toward strengthening the modern theory of optimal tax design. While preference differences have been largely left out of that theory, they are readily apparent in the real world and have long been a staple of broader debates over taxation. Given the potential empirical and normative importance of preference heterogeneity demonstrated in this paper, it seems (at least to us) to be a feature of reality that a convincing theory of optimal taxation ought to include.

Appendix

In this appendix, we prove two results mentioned in the text.

Lemma 1: *If the tax function $T(y)$ is less progressive than the tax function $\hat{T}(y)$, then $T(y)$ is less redistributive than $\hat{T}(y)$.*

Proof.

Suppose on the contrary that $T(y)$ were more redistributive, so that $T(y(n_0)) < \hat{T}(\hat{y}(n_0))$ (note that these taxes on the lowest type agent are likely to be negative). Consider a reform of $\hat{T}(y)$ (not necessarily feasible and incentive compatible) that maintains each agent's earnings $\hat{y}(n)$, but lowers each agent's marginal tax rate to $T'(\hat{y}(n))$ and raises consumption by $\hat{T}(\hat{y}(n_0)) - T(y(n_0)) > 0$. Since earnings are unchanged and consumption is higher, each agent must be made better off by this reform. Now suppose agents are allowed to optimize (while the tax function is adjusted to maintain each type's marginal tax rate). By assumed incentive compatibility of $T(y)$, they will select $y(n)$, and this adjustment cannot decrease utility. Therefore all agents are strictly better off under $T(y)$ than $\hat{T}(y)$, contradicting our assumption that $\hat{T}(y)$ was Pareto efficient. Therefore, $T(y)$ is less redistributive than $\hat{T}(y)$.

■

Next, the proof of Proposition 2 stated the following Lemma:

Lemma 2 *The preference neutral scaling is invariant to the choice of reference wage $\bar{w}$ when the social welfare function takes the utilitarian form in* (9) *and the utility function $U(c, l)$ takes the form in* (10).

Proof.

Given the functional form assumptions in the lemma, $g^{LF}(\theta_i \bar{w}) = U_c\left(y^{LF}(\bar{w}\theta_i), y^{LF}(\bar{w}\theta_i)/(\bar{w}\theta_i)\right)$. Preference neutrality therefore requires that the scaling $\{\alpha_i\}_{i=1}^{I}$ satisfies

$$\frac{U_c(y^{LF}(\bar{w}\theta_i), y^{LF}(\bar{w}\theta_i)/(\bar{w}\theta_i))}{U_c(y^{LF}(\bar{w}\theta_j), y^{LF}(\bar{w}\theta_j)/(\bar{w}\theta_j))} = \frac{\alpha_j}{\alpha_i}, \ \forall i, j. \tag{15}$$

Let ε denote the elasticity of taxable income under the *laissez-faire* tax regime:

$$\varepsilon = \frac{\partial y_i}{\partial(1 - T'(y_i))} \frac{1 - T'(y_i)}{y_i}\bigg|_{T(y)=0} = \frac{\partial y^{LF}(\theta_i w_i)}{\partial w_i} \frac{w_i}{y^{LF}(\theta_i w_i)} = \frac{\partial y^{LF}(\bar{w}\theta_i)}{\partial \bar{w}} \frac{\bar{w}}{y^{LF}(\bar{w}\theta_i)}.$$

Let γ denote the coefficient of relative risk aversion. For the class of utility functions in the statement of the lemma, ε and γ are constant.

To see the effect of varying $\bar{w}$ on the preference neutral scaling, we differentiate (15) with respect to $\bar{w}$, where derivatives are evaluated at $y^{LF}(\bar{w}\theta_i)$ and $y^{LF}(\bar{w}\theta_j)$:

$$\frac{1}{\left(U_c^j\right)^2}\frac{U_c^i \cdot U_c^j}{\bar{w}}\left[\frac{U_{cc}^i y_i^{LF}}{U_c^i}\cdot\frac{\partial y_i^{LF}}{\partial \bar{w}}\cdot\frac{\bar{w}}{y_i^{LF}}-\frac{U_{cc}^j y_j^{LF}}{U_c^j}\cdot\frac{\partial y_j^{LF}}{\partial \bar{w}}\cdot\frac{\bar{w}}{y_j^{LF}}\right].$$

The term in brackets is equal to $-\gamma\varepsilon+\gamma\varepsilon=0$, proving that (15) does not vary with $\bar{w}$, and thus the scaling preference neutral scaling does not vary with $\bar{w}$.

■

References

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

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

[3] 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.

[4] Brett, Craig and John A. Weymark (2003). "Financing Education using Optimal Redistributive Taxation," *Journal of Public Economics.*

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

[6] Diamond, Peter (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88(1) pp. 83-95.

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

[8] Feldstein, Martin and Marian Vaillard Wrobel (1998). "Can State Taxes Redistribute Income?" *Journal of Public Economics.*

[9] 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.*

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

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

[12] Golosov, Mikhael, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2013). "Preference Heterogeneity and Optimal Capital Income Taxation," *Journal of Public Economics.*

[13] 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.

[14] Kahneman, Daniel (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus, and Giroux (see page 401).

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

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

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

[18] Mirrlees, James A. (1976). "Optimal Tax Theory: A Synthesis," *Journal of Public Economics.*

[19] 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.

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

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

[22] Rothschild, Casey and Florian Scheuer (2013). "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics.*

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

[24] Salanié, Bernard. *The Economics of Taxation.* MIT.

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

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

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

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