NBER WORKING PAPER SERIES

GENERALIZED SOCIAL MARGINAL WELFARE WEIGHTS FOR OPTIMAL TAX THEORY

Emmanuel Saez
Stefanie Stantcheva

Working Paper 18835
http://www.nber.org/papers/w18835

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

We thank Steve Coate, Florian Ederer, Marc Fleurbaey, Louis Kaplow, Henrik Kleven, Etienne Lehmann, Ben Lockwood, Thomas Piketty, Larry Samuelson, Maxim Troshkin, Aleh Tsyvinski, Matthew Weinzierl, Nicolas Werquin, and numerous conference participants for useful discussions and comments. We acknowledge financial support from NSF Grant SES-1156240, the MacArthur Foundation, and the Center for Equitable Growth at UC Berkeley. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Generalized Social Marginal Welfare Weights for Optimal Tax Theory
Emmanuel Saez and Stefanie Stantcheva
NBER Working Paper No. 18835
February 2013, Revised October 2014
JEL No. H21

ABSTRACT

This paper proposes a new way to evaluate tax reforms, by aggregating losses and gains of different individuals using "generalized social marginal welfare weights." A tax system is optimal if no budget neutral small reform can increase the weighted sum of (money metric) gains and losses across individuals. Optimum tax formulas take the same form as standard welfarist tax formulas by simply substituting standard marginal social welfare weights with those generalized marginal social welfare weights. Weights directly capture society's concerns for fairness allowing us to cleanly separate individual utilities from social weights. Suitable weights can resolve puzzles of the traditional welfarist approach, as well as unify in an operational way the most prominent alternatives to utilitarianism such as Libertarianism, Rawlsianism, Equality of Opportunity, Poverty alleviation, or Fair Income Taxation. Generalized welfare weights can be specified to, among others, (1) provide a non-confiscatory theory of optimal taxation even absent any behavioral responses, (2) treat differently "deserved income" vs. "undeserved income," (3) distinguish between "deserving transfer beneficiaries" vs. "free loaders," (4) rule out the use of tags unless they create a Pareto improvement. We use a simple online survey to illustrate how to map social preferences of the public into weights.

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

Stefanie Stantcheva
Department of Economics
Littauer Center 232
Harvard University
Cambridge, MA 02138
and NBER
sstantcheva@fas.harvard.edu

1 Introduction

The dominant approach in optimal tax theory is to use the standard *welfarist* framework, in which the government sets taxes and transfers to maximize a social welfare function which is an explicit function of individual utilities and solely of individual utilities. The theory derives optimal tax formulas trading-off equity and efficiency. The optimal tax system is the one around which no further reform is desirable, when the gains and losses across individuals are aggregated using the social marginal welfare weights. These weights represent the value that society puts on providing an additional dollar of consumption to any given agent. The standard approach imposes a specific, stringent, structure on those weights based exclusively and entirely on the individual utility functions and the social welfare function.

In this paper we propose a novel approach that replaces the standard weights by alternative *generalized social marginal welfare weights*. These weights are no longer necessarily derived from an underlying social welfare function but are instead specified so as to directly reflect society's concerns for fairness. Importantly, they can depend on individual and aggregate characteristics, some of which are endogenous to the tax and transfer system. The characteristics which enter the welfare weights determine the dimensions along which society considers redistribution to be fair. These characteristics could be part of individuals' utilities (in the welfarist spirit). Importantly, though, social welfare weights can also depend on individual or aggregate characteristics which do not enter individuals' utilities. Conversely, the welfare weights can omit some characteristics which enter individuals' utility functions, but for which society does not deem it fair to compensate individuals. Naturally, the generalized welfare weights nest the standard welfarist weights, a special case in which private characteristics and social criteria exactly coincide.

We outline a transparent and operational method to solve for optimal tax systems for any set of generalized welfare weights. The first step is to aggregate individual weights. The appropriate level of aggregation depends on the feasibility constraints on the tax system, i.e., what the tax and transfer system can depend on.[1] For instance, if tax system can depend only on income as is standard, the generalized social welfare weights are aggregated at each income level. We then show that the optimum tax formulas take the same form as standard welfarist tax formulas, but with the new generalized social welfare weights replacing standard social welfare weights.[2] The optimal taxes we obtain can be seen as an equilibrium around which no marginal reform

[1] While the weights can depend for instance on individual unobserved characteristics, the tax system cannot.

[2] Other formulas derived for more complex tax systems, for instance with tagging, also hold in the same way replacing standard weights with generalized welfare weights.

is desirable given the generalized social marginal welfare weights. Indeed, Farhi and Werning (2013) and Piketty and Saez (2013b) consider optimal inheritance taxation with heterogeneous tastes for altruism and implicitly use generalized social welfare weights to normalize their Pareto weights. Golosov, Tsyvinski, and Werquin (2013) also express and interpret optimal dynamic tax formulas in terms of marginal social welfare weights.[3]

With non-negative weights, constrained Pareto efficiency applies.[4] While the optimum we obtain is consistent with maximizing a weighted sum of utilities with specific Pareto weights, such Pareto weights cannot be defined a priori and, hence, starting with generalized social welfare weights is required. This new approach, presented in Section 2, has three main advantages.

First, we can address some major weaknesses of the standard approach, thanks to the fact that individual characteristics and social criteria are allowed to be different.[5] With concave individual utilities and the standard *utilitarian* criterion, the marginal welfare weights decrease in after-tax income so that redistributing from high to low incomes is socially valued. This justification for redistribution is opposed both from a normative perspective of justice principles and from a positive perspective of actual social preferences of the public by Nozick (1974) and, more recently, Feldstein (2012), Mankiw (2010, 2013) and Weinzierl (2012). Related, if individuals do not respond to taxes, i.e., if pre-tax incomes are fixed, and individual utilities are concave, then utilitarianism recommends a 100% tax and full redistribution, a point originally made by Edgeworth (1897). In reality, even absent behavioral responses, many–and perhaps most–people would object to complete redistribution and optimal taxation may not be as trivial nor as extreme as standard utilitarian theory would suggest.

Furthermore, views on taxes and redistribution seem largely shaped by views on whether the income generating process is fair and whether individual incomes are deserved or not. The public tends to dislike the redistribution of fairly earned income but is in favor of redistributing income earned unfairly or received through pure luck (Alesina and Angeletos, 2005, Alesina and Giuliano, 2011). Such distinctions are irrelevant under utilitarianism.

In addition, society assesses the value of transfers or the costs of taxes not only based on the actual economic situation of a given person but also based on what this person would have done

[3]Saez (2001, 2002) expressed optimal income tax formulas directly in terms of social welfare weights, hence implicitly using the approach presented here. Golosov and Tsyvinski (2014) survey the dynamic tax literature.

[4]Pareto efficiency has been a major focus of the taxation literature (Battaglini and Coate, 2008). At the optimum, there is no budget neutral small reform that the government could undertake given its informational constraints and that could increase everybody's welfare. Due to the local nature of those weights, our approach can only guarantee local Pareto efficiency, i.e., Pareto efficiency relative to small reforms around the status quo. As we discuss, multiple equilibria can arise in some situations.

[5]The survey papers by Kaplow (2008), Fleurbaey (2008), Piketty and Saez (2013), and Mankiw, Weinzierl, and Yagan (2009) discuss in detail the limitations of the standard optimal tax approach.

absent that transfer or tax. For example, most people would value transfers to those unable to work (the "deserving poor") while fewer people value transfers to those who would stop working because of transfers themselves (the "free loaders") (Will 1993, Larsen 2008, Jeene, van Oorschot, and Uunk, 2011). The distinction between actual and counterfactual outcomes cannot be made in the welfarist approach.

Moreover, in the welfarist approach, optimal taxes should depend not only on income but also on all other observable characteristics which are correlated with intrinsic earning ability, such as height, gender, or race. Yet society seems highly reluctant to make taxes depend on such "tags"(Mankiw and Weinzierl, 2010). The public seems to value horizontal equity, i.e., to find it unfair to tax differently people with the same ability to pay, a concern not easily reconciled with the welfarist approach.

Finally, and thanks to the fact that social preferences may (but need not) be completely independent from individual utilities, we can circumvent better the problem of cardinal utilities. Society needs to take a direct stand on the value of a transfer to any given agent, but not on individuals' cardinal utility representation.

Second, it appears that policymakers do not explicitly posit and maximize a social welfare function based on individual utilities. Actual tax policy debates tend to focus instead on specific tax reforms, starting from a given situation, considering who the winners and losers are, and the broader consequences of the reform on economic activity and tax revenue. Hence, a tax reform approach, that also places the emphasis in the weights attributed to transfers, seems more suitable for policy analysis.[6]

Third, a framework based on the social marginal welfare weights can unify in a tractable and operational framework alternative criteria proposed in the literature, such as Equality of opportunity, Fair income tax, Libertarianism or Rawlsianism, or Poverty alleviation. In particular, we can apply already existing tax formulas, replacing standard weights with the appropriate weights that capture each criteria. This approach can be applied much more generally to the evaluation of any reforms that can be represented by a set of transfers to different agents and for which society features some justice or fairness considerations.

In Section 3, we show how the use of suitable generalized social welfare weights, and in particular, the flexibility to draw a distinction between individual characteristics and social criteria, can help resolve puzzles of the traditional utilitarian approach and account for many existing tax policy debates and tax structures. First, we show that making generalized social

[6] As already explained, the standard approach can also be recast in tax reform terms, but using weights that are restricted by exclusively individual utilities and the social welfare function.

marginal weights depend negatively on net taxes paid, in addition to net disposable income, generates a non-trivial optimal tax theory even absent behavioral responses. Second and related, we show that such generalized social welfare weights depending on net-taxes paid can be micro-founded based on the fact that society prefers to tax income due to luck rather than income earned through hard work. Third, generalized social weights can depend on what individuals would have done absent taxes and transfers. Hence, we can capture the idea that society dislikes marginal transfers toward free loaders who would work absent means-tested transfers. Fourth, our approach can capture horizontal equity concerns. A reasonable criterion is that introducing horizontal inequities is acceptable only if it benefits the group discriminated against. This dramatically limits the scope for using non-income based tags.

In Section 4, we show how the most prominent alternatives to utilitarianism can be re-cast within our theory, i.e., we can derive the generalized social welfare weights implied by those alternative theories.[7] First, the Rawlsian criterion concentrates social welfare weights solely on the most disadvantaged in society. Second, the Libertarian criterion concentrates weights on those who contribute more in taxes than they receive in transfers or public goods, following the benefits principle for taxation. Third, the equality of opportunity principle developed by Roemer (1998) and Roemer et al. (2003) concentrates weights uniformly on those coming from a disadvantaged background as, conditional on earnings, they have more merit (have worked harder) than those coming from an advantaged background. As the likelihood of coming from a disadvantaged background decreases with income, social weights decrease with income for a reason completely orthogonal to the decreasing marginal utility of income in utilitarianism. It also provides a rationale for less progressive taxes when there is high social or intergenerational mobility. Finally, poverty alleviation criteria that respect the Pareto principle can also be captured by social welfare weights concentrated on those below the poverty threshold.

Finally, Section 5 shows that generalized welfare weights could also be derived empirically, by estimating actual social preferences of the public, leading to a positive theory of taxation. There is indeed a small body of work trying to uncover perceptions of the public about various tax policies. These approaches either start from the existing tax and transfers system and reverse-engineer it to obtain the underlying social preferences (Christiansen and Jansen 1978, Bourguignon and Spadaro 2012, Zoutman, Jacobs, and Jongen 2012) or directly elicit preferences on various social issues in surveys.[8] Using a simple online survey with over 1000 participants,

[7] To do this, we consider the case of optimal income taxation where we use Saez (2001) formulas expressed in terms of social marginal welfare weights and valid for heterogeneous populations (Lehmann and Jacquet, 2013).

[8] See Yaari and Bar-Hillel (1984), Frohlich and Oppenheimer (1992), Cowell and Shokkaert (2001), Fong

we illustrate how public preferences can be mapped into generalized social marginal welfare weights. Our results confirm that public views on redistribution are inconsistent with standard utilitarianism. Section 6 concludes. All proofs are in the Online Appendix.

2 Outline of our Approach

We outline our approach using income taxation.[9] Consider a population with a continuum of individuals indexed by i. Population size is normalized to one. Individual i derives utility from consumption c_i and incurs disutility from earning income z_i, with a utility representation:

$$u_i = u(c_i - v(z_i; x_i^u, x_i^b))$$

where x_i^u and x_i^b are vectors of characteristics, u increasing and v increasing in z. The functions u and v are common to all individuals. u_i is a cardinal utility representation for individual i as viewed by the Planner (see below). Contrary to the standard approach, social welfare weights need not depend on u_i in any way and a cardinal utility representation is not necessary.

Important characteristics considered throughout the paper are, first, a person's productivity per unit of effort $w_i \equiv z_i/l_i$ where l_i is labor supply. w_i is distributed in the population with a density $f(w)$ on $[w_{\min}, w_{\max}]$. Second, we consider the cost of work θ_i that affects the disutility from producing any unit of effort, distributed according to a distribution $p(\theta)$ on $[\theta_{\min}, \theta_{\max}]$. For instance, the disutility from labor could take the form $\theta_i \tilde{v}(z_i/w_i)$, so that any unit of effort is more costly for high θ_i agents. More precisely, x^u are characteristics that exclusively enter the utility function, while x^b will be characteristics that will also affect the generalized social welfare weights introduced below. It is entirely possible to consider a more general utility with $u_i = u(x_i^c \cdot c_i - v(z; x_i^u, x_i^b))$ where x_i^c would be a shifter parameter for the marginal utility from consumption. For simplicity, we abstract from his heterogeneity as none of our examples, except a short discussion in Section 3.1 requires it.

The government sets an income tax $T(z)$ as a function of earnings only so that $c_i = z_i - T(z_i)$.[10] Individual i chooses z_i to maximize $u(z_i - T(z_i) - v(z_i; x_i^u, x_i^b))$.

(2001), Devooght and Schokkaert (2003), Engelmann and Strobel (2004), Ackert, Martinez-Vazquez, and Rider (2007), Gaertner and Schokkaert (2012), Kuziemko, Norton, Saez, and Stantcheva (2013), Weinzierl (2012). Our focus on tax reform and on (local) marginal welfare weights might make it much easier to elicit social preferences than if trying to calibrate a global objective function.

[9] The same approach can be applied to other forms of taxation or types of policies.

[10] We consider in Section 3.4 the situation where the tax system also depends on other observable characteristics of the individual, called "tags".

2.1 Standard welfarist approach framed as a tax reform approach

In the standard welfarist approach, the government chooses the tax schedule $T(z)$ to maximize a social welfare objective that is a sole function of individual utilities.

$$SWF = \int_i \omega_i \cdot u_i, \tag{1}$$

subject to: (1) the aggregate budget constraint $\int_i T(z_i) \geq E$ where E is exogenous (non-transfer related) government spending, (2) the fact that individual earnings z_i respond to taxes ("incentive compatibility constraints"). To economize on notation, we denote sums over the population simply by $\int_i$ without specifying the measure.

The non-negative weights ω_i are given exogenously and are not allowed to depend directly on the tax system or on exogenous characteristics of people. Kaplow and Shavell (2001) make the important point that including any other endogenous elements in the weights can lead to Pareto dominated outcomes in some circumstances.

Definition 1 *The standard social marginal welfare weight of individual i is: $g_i \equiv \omega_i \cdot u_{ci}$, with $u_{ci} \equiv u'(c_i - v(z_i; x_i^u, x_i^b))$ the marginal utility of consumption.*[11]

Using the standard envelope argument from the individual's optimization of z_i, a small tax reform $dT(z)$ changes utility $u_i = u(z_i - T(z_i) - v(z_i; x_i^u, x_i^b))$ by $du_i = -u_{ci} \cdot dT(z_i)$. Hence, the behavioral response dz_i can be ignored and $dT(z_i)$ measures the money-metric welfare impact of the tax reform on individual i. The net effect on social welfare is $-\int_i \omega_i u_{ci} \cdot dT(z_i) = -\int_i g_i dT(z_i)$. Because individuals adjust their earnings z_i by dz_i following the reform, the change in taxes paid by person i is $dT(z_i) + T'(z_i)dz_i$, where $T'(z_i)dz_i$ is the fiscal change due to the behavioral response dz_i.

Definition 2 *A reform $dT(z)$ is budget neutral if and only if $\int_i [dT(z_i) + T'(z_i)dz_i] = 0$.*

Definition 3 ***Tax reform desirability criterion (standard approach).*** *A small budget neutral tax reform $dT(z)$ around the current tax system $T(z)$ is desirable if and only if $\int_i g_i dT(z_i) < 0$, with $g_i = \omega_i \cdot u_{ci}$ the social marginal welfare weight on i evaluated at $T(z)$.*

Proposition 1 ***Optimal tax criterion (standard approach).*** *If a tax system $T(z)$ is optimal, then for any budget neutral, small tax reform $dT(z)$, $\int_i g_i dT(z_i) = 0$ with g_i the standard social marginal welfare weight on individual i.*

[11] Note that if the social welfare function is specified as $SWF = \int_i G(u_i)$ where G is concave and increasing in individual utilities, then the social welfare weight of i is: $g_i = G'(u_i) \cdot u_{ci}$.

Proof: If $\int_i g_i dT(z_i) < 0$, the reform would increase social welfare. If $\int_i g_i dT(z_i) > 0$ then $-dT(z)$ increases social welfare and is also budget neutral. ■

Knowing behavioral responses is necessary to assess whether a tax reform $dT(z)$ is budget neutral. Once this is known, assessing the welfare effects of $dT(z)$ just requires evaluating the mechanical effects of the reform on each individual (i.e., ignoring behavioral responses dz_i) and weighting them by the social marginal welfare weights g_i from Definition 1.[12]

2.2 The Optimal tax formula and its inversion

The reform approach in Proposition 1 yields a clear optimal tax formula that is particularly simple in the case of no income effects on labor supply that we consider here (Diamond, 1998). See Appendix A.1 for a short derivation and Piketty and Saez (2013) for details.[13] Let $H(z)$ be the cumulative earnings distribution function and $h(z)$ the earnings density.

Definition 4 *Let $\bar{G}(z)$ be the (relative) average social marginal welfare weight for individuals who earn more than z:*

$$\bar{G}(z) \equiv \frac{\int_{\{i:z_i \geq z\}} g_i}{Prob(z_i \geq z) \cdot \int_i g_i} \tag{2}$$

Let $\bar{g}(z)$ be the corresponding average social marginal welfare weight at earnings level z, with $\bar{G}(z)[1 - H(z)] = \int_z^\infty \bar{g}(z')dH(z')$, or, equivalently:

$$\bar{g}(z) = -\frac{1}{h(z)} \frac{d(\bar{G}(z) \cdot [1 - H(z)])}{dz} \tag{3}$$

Social marginal welfare weights average to one in the population: $\bar{G}(0) = \int_0^\infty \bar{g}(z)dH(z) = 1$.

Result 1 *The optimal marginal tax at income level z is given by:*

$$T'(z) = \frac{1 - \bar{G}(z)}{1 - \bar{G}(z) + \alpha(z) \cdot e(z)} \tag{4}$$

with $e(z)$ the average elasticity of earnings z_i with respect to the retention rate $1 - T'$ for individuals earning $z_i = z$, $\alpha(z)$ the local Pareto parameter defined as $zh(z)/[1 - H(z)]$.[14]

[12] The tax reform approach only provides necessary first order conditions. This does not guarantee that the global maximum has been reached, i.e., we can be at a local extremum including possibly a minimum (see e.g., Guesnerie, 1995 for a detailed discussion).

[13] Saez (2001) informally derived an optimal income tax formula that generalizes the formulas of Mirrlees (1971) to situations with heterogeneous populations, where individuals differ not only in skills but also possibly in preferences. Lehman and Jacquet (2013) provide a fully rigorous proof of the formula with heterogeneous populations and give conditions under which it applies. We always assume here that these conditions hold.

[14] The local Pareto parameter $\alpha(z)$ is constant and equal to the Pareto parameter for Paretian distributions.

Any budget balanced tax schedule with $T'(z) < 1$ can be inverted to obtain the corresponding social welfare weights (Werning, 2007, Hendren, 2013).

Proposition 2 *If $T'(z)$ exists and $T'(z) < 1$ for all z, there is an unique $\bar{G}(z) < 1 + \alpha(z) \cdot e(z)$ defined by $\bar{G}(z) = [1 - T'(z)(1 + \alpha(z) \cdot e(z))]/[1 - T'(z)]$ that satisfies the optimal formula (4). The corresponding $\bar{g}(z)$ is then given by formula (3).*

There is no guarantee that the implied weights are non-negative, i.e., that the initial tax system is Pareto efficient. The weights $\bar{g}(z)$ may also well be discontinuous or Dirac functions.

We can similarly express the optimal *linear* tax τ as a function of the welfare weights:[15]

Result 2 *The optimal linear income tax is:*

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + e} \qquad \textit{with} \qquad \bar{g} \equiv \frac{\int_i g_i \cdot z_i}{\int_i g_i \cdot \int_i z_i} \tag{5}$$

and e the elasticity of aggregate income $\int_i z_i$ with respect to the retention rate $(1 - \tau)$.

In the linear tax case, $\bar{g}$ can be interpreted as the average g_i weighted by income z_i (relative to the population average g_i) or symmetrically as the average z_i weighted by g_i (relative to $\int_i z_i$).

2.3 Generalized social welfare weights approach

We propose a novel theory of taxation that starts directly from the social welfare weights. For any individual, we can define a *generalized social marginal welfare weight* g_i which measures how much society values the marginal consumption of individual i.

Definition 5 *The generalized social marginal welfare weight on individual i is:*

$$g_i = g(c_i, z_i; x_i^s, x_i^b) \tag{6}$$

where g is a function, x_i^s is a vector of characteristics which only affect the social welfare weight, while x_i^b is a vector of characteristics which also affect utility.

Naturally, the generalized weights are only defined up to a multiplicative constant as they measure only the *relative* value of consumption of individual i. Importantly, they are allowed

To be precise, when defining $\alpha(z)$, $h(z)$ is defined as the virtual density that would hold at z if the income tax system were linearized at z. See Saez (2001) and Piketty and Saez (2013) for details.

[15] See again Appendix A.1 for a short derivation and Piketty and Saez (2013) for details.

to depend on individual characteristics x_i^s and x_i^b. These characteristics may be unobservable to the government or it may be impossible or unacceptable to condition the tax system on them. They nevertheless enter the social welfare weights because they affect how deserving a person is deemed by society. For instance, x_i^s might include family background (see our treatment of "Equality of Opportunity" in section 4.2), a characteristic that typically does not affect one's taxes directly but affects perceptions of deservedness.

There is an important conceptual distinction between the sets of characteristics x^b, x^u, and x^s. Characteristics which enter the social welfare weight are dimensions that society considers *fair to redistribute across* and to compensate for. For instance, if the disutility of work θ_i is judged to be due mostly to differences in health status or disability, then it might be fair to include it in the social welfare weight (see the "Free Loaders" example in section 3.3). On the other hand, if differences in disutility for work are mostly based on varying degrees of laziness, then it might be considered socially unfair to compensate people for these (see the "Fair Income Tax" example in section 4.4). These value judgments are directly embodied in the specification of the social welfare weights.

Based on these weights, we can define an equilibrium again using the tax reform approach:

Definition 6 ***Tax reform desirability criterion (generalized approach).*** *A small budget neutral tax reform $dT(z)$ around the current tax system $T(z)$ is desirable if and only if $\int_i g_i dT(z_i) < 0$, with g_i the generalized social marginal welfare weight on individual i evaluated at $(c_i = z_i - T(z_i), z_i, x_i^s, x_i^b)$.*

Proposition 3 ***Optimal tax criterion (generalized approach).*** *A tax system $T(z)$ is optimal if and only if, for any budget neutral, small tax reform $dT(z)$, $\int_i g_i dT(z_i) = 0$, with g_i the generalized social marginal welfare weight on individual i evaluated at $(z_i - T(z_i), z_i, x_i^s, x_i^b)$.*

As in the standard theory (see e.g. Kleven and Kreiner 2006, Chetty 2009, Hendren, 2013), the tax reform approach requires knowing only the weights g_i and the behavioral responses around the current system while the optimal tax criterion requires knowing the weights g_i and the behavioral responses *at the optimum*. Hence, the tax reform is more readily applicable.

From individual weights to applicable weights in the tax formula. The weights on each individual g_i are not immediately applicable. As noted, they can depend on unobservable characteristics or elements that the tax system cannot condition upon. They merely embody society's judgment of fairness, without criteria such as observability or feasibility. To apply

the weights to the evaluation of tax systems, the individual weights need to be "aggregated" up to only those characteristics that the tax system can be conditioned on. For instance, for a sophisticated tax system that can depend on both income z and some characteristic x^b, the weights would have to be aggregated for each pair (z, x^b). For a tax system purely based on income, $T(z)$, the weights need to be aggregated at each income level z. This gives $\bar{G}(z)$ and $\bar{g}(z)$ as in formulas (2) and (3) where g_i are the generalized social welfare weights. We can then immediately use formula (4) with these generalized weights, aggregated at each income level.[16]

Advantages of the generalized approach. Our approach with generalized weights has three advantages relative to the standard welfarist approach, as well as to the alternatives to welfarism proposed in the literature.

First, it allows us to immediately exploit the previously derived tax formulas such as equation (4) once they are written in terms of the welfare weights at each income level. It nests the standard approach. Furthermore, it allows to unify alternative approaches to welfarism in one tractable framework (see Section 4).

Second, it can easily ensure that any tax optimum is constrained Pareto efficient as long as the generalized weights g_i are all non-negative. To see this, the optimum is equivalent to maximizing the linear social welfare function $SWF = \int_i \omega_i \cdot u_i$ with *Pareto weights* $\omega_i = g_i/u_{ci} \geq 0$ where g_i and u_{ci} are evaluated at the optimum allocation (i.e., are taken as fixed in the maximization of SWF).[17] Hence, our approach can be reverse-engineered to obtain a set of Pareto weights ω_i and a corresponding standard social welfare function $\int_i \omega_i \cdot u_i$. However, in practice as we shall see, it is impossible to posit the correct weights ω_i without *first* having solved for the optimum using our approach that starts with the social marginal weights g_i.[18] Related to this, our approach does not use a pre-specified social objective to be maximized. Hence, our optimum is really an equilibrium around which no small budget neutral reform is desirable when weighting gains and losses using the weights g_i.

Third and most importantly, as we will outline throughout the paper, our approach grants great flexibility in the choice of the welfare weights g_i. The fact that social welfare weights can depend on characteristics outside of individuals' utilities, as well as ignore characteristics from

[16] This can be done either by estimating the distribution of the characteristics (x_i^s, x_i^b) at each income level z in the data (which does not require any assumption on individual utilities) or by inferring it from behavior.

[17] This assumes that the first order condition characterizes the optimum. Absent this assumption, our generalized optimum may only be a local constrained Pareto optimum. We come back to this point below.

[18] A closely related point has also been made in Fleurbaey and Maniquet (2013) who show that their fair income tax social objective maximization can also be obtained as the maximization of a weighted sum of utilities but the weights "would have to be computed for each new problem, that is, as a function of the set of allocations among which the choice has to be made,..., and could only be computed after the optimal allocation is identified."

individuals' utilities allows us to incorporate elements that matter in actual tax policy debates and yet cannot be captured with the standard welfarist approach.

3 Resolving Puzzles of the Welfarist Approach

In this section, we show how the use of suitable generalized social marginal welfare weights can resolve the main puzzles that arise in the welfarist approach to optimal tax theory. Table 1 summarizes these results by contrasting actual tax practice (column 1), the standard welfarist approach (column 2), and our generalized social marginal welfare weights approach (column 3) in various situations. In each situation, column 3 indicates what property of social marginal welfare weights is required to make this approach fit with actual tax policy practice.

3.1 Optimal Tax Theory with Fixed Incomes

We start with the simple case in which pre-tax incomes are completely inelastic to taxes and transfers. This puts the focus solely on the redistributive issues. It is a useful introduction to our approach, especially as contrasted with the standard welfarist approach. We specialize our general framework with a disutility of work $v(z; z_i) = 0$ if $z \leq z_i$ and $v(z; z_i) = \infty$ if $z > z_i$. Thus, z_i is an exogenous characteristic of individual i, contained in x_i^u, and choosing $z = z_i$ is always optimal for the individual, so that the distribution of incomes $H(z)$ is exogenous to the tax system. In equilibrium, utility is $u_i = u(c_i)$. We review first the standard utilitarian setting.

Standard utilitarian approach. The government chooses $T(z)$ to maximize the standard social welfare function in (1) subject to the resource constraint $\int T(z)dH(z) \geq 0$ (with multiplier p). A point-wise maximization with respect to $T(z)$ yields $u'(z - T(z)) = p$ so that $c = z - T(z)$ is constant across z. Hence, utilitarianism with inelastic earnings and concave individual utility functions leads to complete redistribution of incomes. The government taxes 100% of earnings and redistributes income equally across individuals (Edgeworth, 1897). The optimum is such that all standard marginal welfare weights $g_i = u_{ci}$ are equalized across individuals.

This simple case highlights three of the drawbacks of utilitarianism. First, complete redistribution seems too strong a result. In reality, even absent behavioral responses, many and perhaps even most people would still object to 100% taxation on the grounds that it is unfair to fully confiscate individual incomes. Second, the outcome is extremely sensitive to the specification of individual utilities, as linear utility calls for no taxes at all, while introducing just a bit of concavity leads to complete redistribution. Third, the utilitarian approach cannot handle well

heterogeneity in individual utility functions. This is known as the problem of inter-personal utility comparisons. Take the more general formulation with fixed incomes and heterogeneity in the marginal value of consumption $u_i = u(x_i^c \cdot c_i)$. The optimum is such that $x_i^c u'(x_i^c \cdot c_i)$ are equal for all i. Hence, consumption is no longer necessarily equal across individuals and is higher for individuals more able to enjoy consumption. In reality, society would be reluctant to redistribute based on preferences, which we confirm with our online survey in Section 5.[19]

Generalized Social Marginal Welfare Weights. The simplest way to illustrate the power of our approach with fixed incomes is to use the generalized weights as defined in (6) without using any additional characteristics.

Definition 7 Simple generalized weights: *Let* $g_i = g(c_i, z_i) = \tilde{g}(c_i, z_i - c_i)$ *with* $\tilde{g}_c \equiv \frac{\partial \tilde{g}}{\partial c} \leq 0$, $\tilde{g}_{z-c} \equiv \frac{\partial \tilde{g}}{\partial (z-c)}|_c \geq 0$.[20] *There are two polar cases of interest:*
i) Utilitarian weights: $g_i = g(c_i, z_i) = \tilde{g}(c_i)$ *for all* z_i*, with* $\tilde{g}(\cdot)$ *decreasing.*
ii) Libertarian weights: $g_i = g(c_i, z_i) = \tilde{g}(z_i - c_i)$ *with* $\tilde{g}(\cdot)$ *increasing.*

Weights depend not only negatively on c but also positively on net taxes paid $z - c$. $\tilde{g}_c \leq 0$ reflects the fact that society values additional consumption less (when keeping taxes paid constant). This captures the old notion of "ability to pay" as under utilitarianism with a concave utility of consumption. On the other hand, a higher tax burden $(z - c)$ increases the weight, since taxpayers contribute more to society and are more deserving of additional consumption. Another interpretation is that individuals are in principle entitled to their income and hence become more deserving as the government taxes away their income.[21]

The optimal tax system, according to Proposition 3 is such that no reform can increase social welfare at the margin, where transfers are evaluated using the g weights. Since, $T(z) = z - c$, $\tilde{g}(c, z - c) = \tilde{g}(z - T(z), T(z))$. With no behavioral responses, the optimal rule is simple: social welfare weights $\tilde{g}(z - T(z), T(z))$ need to be equalized across all incomes z. Intuitively, if we had non equalized weights with $\tilde{g}(z_1 - T(z_1), z_1) > \tilde{g}(z_2 - T(z_2), z_2)$, transferring a dollar from those earning z_2 toward those earning z_1 (by adjusting $T(z_1)$ and $T(z_2)$ correspondingly and in a budget balanced manner) would be desirable (the formal proof is in Appendix A.2).

Proposition 4 *The optimal tax schedule with no behavioral responses is characterized by:*

$$T'(z) = \frac{1}{1 - \tilde{g}_{z-c}/\tilde{g}_c} \quad \textit{and} \quad 0 \leq T'(z) \leq 1. \tag{7}$$

[19] Redistribution based on marginal utility is socially acceptable if there are objective reasons a person has higher needs, such as having a medical condition requiring high expenses, or a large family with many dependents.
[20] In our general notation, $x_i^u = z_i$, for all i, and x^b and x^s are empty.
[21] We assume away government funded public goods in our setup for simplicity.

Corollary 1 *In the standard utilitarian case, $T'(z) \equiv 1$. In the libertarian case, $T'(z) \equiv 0$.*

We present in Section 5 results from a survey asking subjects to rank taxpayers with various incomes and tax burdens in terms of deservedness of a tax break. As stipulated here, respondents put weight on both disposable income and gross income, showing that social preferences are in between the polar utilitarian and libertarian cases. Such data can be used to recover social preferences $g(c, z)$. For instance, the specification $g(c, z) = \tilde{g}(c - \alpha(z - c)) = \tilde{g}(z - (1 + \alpha)T(z))$ with $\tilde{g}$ decreasing and where α is a constant parameter delivers an optimal tax with a constant marginal tax rate $T'(z) = 1/(1 + \alpha)$ and will be calibrated in Section 5.

3.2 Luck Income vs. Deserved Income

A widely held view is that it is fairer to tax income due to "luck" than income earned through effort and that it is fairer to insure against income losses beyond individuals' control.[22] Our framework captures in a tractable way such social preferences, which differentiate income streams according to their source.[23] These preferences can also provide a micro-foundation for generalized social welfare weights $\tilde{g}(c, z - c)$ increasing in $T = z - c$, as presented in Section 3.1 above.

Suppose there are two sources of income: y^d is deserved income, due to one's own effort, and y^l is luck income, due purely to one's luck. We assume first that y^l and y^d are exogenously distributed in the population and independent of taxes (we consider below the case with elastic effort). Total income is $z = y^d + y^l$. Let us denote by Ey^l average luck income in the economy.

Consider a society with the following preferences for redistribution: Ideally, all luck income y^l should be fully redistributed, but individuals are fully entitled to their deserved income y^d. These social preferences can be captured by the following binary set of weights:

$$g_i = 1(y_i^l - Ey^l \leq z_i - c_i) \tag{8}$$

In our notation, $x_i^s = (y_i^l, Ey^l)$, with Ey^l being an aggregate characteristic common to all agents. A person is "deserving" and has a weight of one if its tax confiscates more than the excess of her luck income relative to average luck income. Otherwise, the person receives a zero weight.[24]

[22] See e.g., Fong (2001) and Devooght and Shokkaert (2003) for how the notion of control over one's income is crucial to identify what is deserved income and Cowell and Shokkaert (2001) for how perceptions of risk and luck inform redistributive preferences.

[23] The problem of luck vs. deserved income is also discussed in Fleurbaey (2008), chapter 3.

[24] In this illustration, we have considered the special case of binary individual weights. More generally, we could specify weights in a continuous fashion based on the difference between $y_i^l - Ey^l$ and $z_i - c_i$. Such alternative weights would also provide a micro-foundation for the function $\tilde{g}(c, z - c)$.

Observable luck income: Suppose first that the government is able to observe luck income and condition the tax system on it, with $T_i = T(z_i, y_i^l)$. In this case, as discussed in Section 2.3, it is necessary to aggregate the individual g_i weights in (8) at each (z, y^l) pair. The aggregated weights are given by: $\bar{g}(z, y^l) = 1(z - T(z, y^l) \leq z - y^l + Ey^l)$, where Ey^l is a known constant, independent of the tax system. Hence, the equilibrium is to, first, ensure everybody's luck income is just equal to Ey^l with $T(z, y^l) = y^l - Ey^l + T(z)$ where $T(z)$ is now a standard income tax set according to formula (7), which leads to $T(z) = 0$, as society does not want to redistribute deserved income. A real-world example of luck income is health care. Health costs are effectively negative luck income and the desire to compensate people for them leads to universal health insurance in all advanced economies.

Unobservable luck income. With unobservable luck income, we make the assumption that any change in total income is partially driven by luck income and partially by deserved income. **Assumption:** For any dz_i, we have either: i) $0 < dy_i^l < dz_i$ or ii) $0 > dy_i^l > dz_i$.

If luck income is not observable, taxes can only depend on total income, with $T_i = T(z_i)$. This model can provide a micro-foundation for the generalized weights $\tilde{g}(c, z - c)$ introduced in Definition 7. If we aggregate the individual weights at each (c, z), we obtain $\tilde{g}(c, z - c) = Prob(y_i^l - Ey^l \leq z_i - c_i | c_i = c, z_i = z)$. Increasing $(z - c)$ at c constant means increasing z. Given that the increase in income is partially driven by luck income and partially by deserved income, y^l increases by less than $z - c$ and $\tilde{g}(c, z - c)$ increases. On the other hand, increasing c while keeping $z - c$ constant means that income increases in the same proportion with $dz = dc > 0$. Given the assumption, again, y^l increases as well, so that $\tilde{g}(c, z - c)$ decreases. Hence, despite the absence of behavioral effects here, the social weights depend positively on $z - c$, even controlling for c. As in Proposition 4, the optimal tax system $T(z)$ equalizes $\tilde{g}(z - T(z), T(z))$ across all z. The presence of indistinguishable deserved income and luck income is enough to generate a non-trivial theory of optimal taxation, even in the absence of behavioral responses.

Beliefs about what constitutes luck income versus deserved income will naturally play a large role in the level of optimal redistribution with two polar cases. If all income is deserved, as libertarians believe in a well-functioning free market economy, the optimal tax is zero. Conversely, if all income were due to luck, the optimal tax is 100% redistribution. If social beliefs are such that high incomes are primarily due to luck while lower incomes are deserved, then the optimal tax system will be progressive.

Behavioral responses and multiple equilibria. If we assume that deserved income responds to taxes and transfers (for example through labor supply responses), while luck income does

not, we can obtain multiple equilibria. Individuals are allowed to differ in their productivity. Utility is $u_i = u(c_i - v(z_i - y_i^l, w_i))$ where w_i is productivity. In this case, $x_i^u = w_i$, $x_i^b = y_i^l$ and $x_i^s = Ey^l$, again common to all agents. The rest of the notation is as in Result 2.

Proposition 5 *With behavioral responses and a linear tax system with rate τ, there are multiple equilibria if the elasticity e of deserved income with respect to $(1-\tau)$ is sufficiently elastic at low tax rates (namely, $e > 1$ at $\tau = 0.1$) and sufficiently inelastic at high tax rates (namely, at $\tau = 0.9$, $e < \bar{g}_0/9$ where $\bar{g}_0$ is as defined in formula (5) evaluated at $\tau = 0$).*

Economies with social preferences favoring hard-earned income over luck income can end up in two possible situations. In the low tax equilibrium, people work hard, luck income makes up a small portion of total income and hence, in a self-fulfilling manner, social preferences tend to favor low taxes. In the alternative equilibrium, high taxes lead people to work less, which implies that luck income represents a larger fraction of total income. This in turn pushes social preferences to favor higher taxes, to redistribute away that unfair luck income (itself favored by the high taxes in the first place). Thus, our framework can encompass the important multiple equilibria outcomes of Alesina and Angeletos (2005) without departing as drastically from optimal income tax techniques.[25]

Note that although each of the equilibria is locally Pareto efficient, the low tax can well Pareto dominate the high tax equilibrium. The tax reform approach is inherently local.

3.3 Transfers and Free Loaders

The public policy debate often focuses on whether non-workers are deserving of support or not. Transfer beneficiaries are deemed deserving if they are truly unable to work, that is, if absent any transfers, they would still not work and live in great poverty without resources. Conversely, they are considered non-deserving, or "free loaders" if they could work and would do so absent more generous transfers. The presence of such "free loaders", perceived to take undue advantage of a generous transfer system, is precisely why many oppose welfare (see e.g., Ellwood (1988), Ellwood and Bane (1996)). It is also the reason why many welfare programs try to target populations which are deemed more vulnerable and less prone to taking advantage

[25] In Alesina and Angeletos (2005), the preferences of the agents are directly specified so as to include a taste for "fairness," while social preferences are standard. We leave individual preferences unaffected and load the fairness concern exclusively onto the social preferences. We find this more appealing because, first, this allows a separation between private and social preferences that do not always coincide in reality and, second, because it leaves individual preferences fully standard.

of the system. Historically, disabled people, widows, and later on single parents have been most likely to receive support from the government.[26] Our online survey from Section 5 confirms that people have strong views on who, among those out-of-work, is deserving of support. We consider a basic model to explain why the standard approach cannot tackle this issue and how generalized weights can be used to capture the concept of free-loaders.

Model. Starting from our general model, assume that individuals can either work and earn a uniform wage w, or not work and earn zero. Because the wage is common across all agents, labor l is proportional to income z and we can rewrite our general utility (and the welfare weights below) as a function of l instead of z: $u_i = u(c_i - v(l_i; \theta_i))$. We take the special functional form: $u(c_l - \theta \cdot l)$ where $l \in \{0, 1\}$ takes the value 1 if an individual works and 0 otherwise. Consumption c_l is equal to c_0 if an individual is out of work and to $c_1 = w \cdot (1 - \tau) + c_0$ if she works, where τ is the earnings tax rate.[27] Taxes fund the transfer c_0. The cost of work θ is distributed according to a cdf $P(\theta)$ and is private information. An individual with cost of work θ works if and only if $\theta \leq c_1 - c_0 = (1 - \tau) \cdot w$. Hence, the fraction of people working is $P(w(1 - \tau))$. Let e be the elasticity of aggregate earnings $w \cdot P(w(1 - \tau))$ with respect to the retention rate $(1 - \tau)$.

This model is a special case of the optimal linear tax model discussed in Section 2. Hence, we can immediately apply formula (5) so that $\tau = (1-\bar{g})/(1-\bar{g}+e)$. In this model, as non-workers have $z_i = 0$ and workers have $z_i \equiv w$, we have $\bar{g} = \int_i g_i z_i / (\int_i g_i \cdot \int_i z_i) = \bar{g}_1 / [P \cdot \bar{g}_1 + (1 - P) \cdot \bar{g}_0]$ where $\bar{g}_1$ is the average g_i on workers, and $\bar{g}_0$ is the average g_i on non-workers.

Standard utilitarian approach. Under the utilitarian objective, we have $g_i = u'(c_0)$ for all non-workers so that $\bar{g}_0 = u'(c_0)$.[28] Hence, the utilitarian approach cannot distinguish between the deserving poor and free loaders. The standard social welfare weight placed on non-workers depends only on c_0 and is completely independent of whether the person would have worked absent taxes and transfers: The standard approach cannot take into account counterfactuals.

Generalized social welfare weights. The generalized weights allow us to treat differently the deserving poor from the free loaders. Formally, let us define the deserving poor as those

[26] The origins of the US welfare system since 1935, starting with the Aid to Families with Dependent Children (AFDC) and continuing with the Temporary Assistance to Needy Families (TANF) federal assistance programs highlights exactly that logic. An alternative explanation is that stigmatizing welfare recipients can be desirable under some assumptions (Besley et al. 1988, Besley and Coate, 1992b).

[27] As there are only two earnings outcomes, 0 and w, this tax system is fully general.

[28] For workers, $g_i = u(c_1 - \theta_i)$ so that $\bar{g}_1 = \int_{\theta \leq (1-\tau)w} u'(c_0 + (1 - \tau)w - \theta) dP(\theta)/P$.

with $\theta > w$, (who would not work, even absent any transfer), and the free loaders as those with $w \geq \theta > w \cdot (1-\tau)$ (who do not work because of the existence of transfers). Denoting by $P_0 = P(w)$ the fraction working when $\tau = 0$, there are $P(w(1-\tau))$ workers, $1 - P_0$ deserving poor, and $P_0 - P(w(1-\tau))$ free loaders.

Let us assume that society sets social marginal welfare weights as follows. Workers obtain a standard weight $g_i = u'(c_1 - \theta_i)$ if $l_i = 1$. The deserving poor obtain a standard weight $g_i = u'(c_0)$ if $l_i = 0$ and $\theta_i \geq w$. Finally, the free-loaders obtain a weight $g_i = 0$ if $l_i = 0$ and $\theta_i < w$. Given that the wage is common across all agents, the weight function is $g_i = g(c, z/w; \theta_i, w)$ where w is an aggregate characteristic, common to all agents. In our general notation, $x_i^b = (\theta_i, w)$. Note that the cost of work θ enters into the social welfare weights. This means that it is considered fair to compensate people for their differential cost of work.[29]

With such weights, we have $\bar{g}_0 = u'(c_0) \cdot (1-P_0)/(1-P)$ as only a fraction $(1-P_0)/(1-P)$ of the non-workers are deserving. Hence, $\bar{g}_0$ is lower relative to the utilitarian case. $\bar{g}_1$ is unchanged relative to the utilitarian case. Therefore, $\bar{g} = \bar{g}_1/[P \cdot \bar{g}_1 + (1-P) \cdot \bar{g}_0]$ is now *higher* than in the utilitarian case and the optimal tax rate $\tau = (1-\bar{g})/(1-\bar{g}+e)$ is correspondingly lower (keeping e and P constant). As expected, the presence of free loaders reduces the optimal tax rate τ relative to the standard case. In the extreme case in which all non-workers are free-loaders, the optimal transfer (and hence the taxes financing it) is zero. This corresponds to the (extreme) view that all unemployment is created by an overly generous welfare system. As long as there are some deserving poor though, taxes and transfers will be positive.

For a given τ, when e is larger, $(1-P_0)/(1-P)$ is smaller as more people stop working to become free loaders. Hence a higher e reduces the optimal tax rate $\tau = (1-\bar{g})/(1-\bar{g}+e)$ not only through the standard efficiency effect e but also through the social welfare weight channel $\bar{g}$ as it negatively affects society's view on how deserving the poor are.

This example also illustrates that *ex post*, it is possible to find suitable Pareto weights for the welfarist approach that can rationalize the tax rate τ^* obtained with generalized welfare weights. In this case, maximizing $\int_\theta \omega(\theta) \cdot u \cdot dP(\theta)$ with $\omega(\theta) = 1$ for $\theta \leq w \cdot (1-\tau^*)$ (workers) and $\theta \geq w$ (deserving poor) and $\omega(\theta) = 0$ for $w \cdot (1-\tau^*) < \theta < w$ (free loaders). However, the Pareto weights ω depend on the equilibrium tax rate τ^* and hence, cannot be specified *ex ante*.

Application 1: Desirability of in-work benefits. As shown in Piketty and Saez (2013), in the intensive labor supply model, the optimal marginal tax rate at the bottom with zero

[29] If θ were interpreted as pure laziness for instance, then this might not be the case (as in Fleurbaey and Maniquet, 2008, also covered below).

earnings takes the simple form $T'(0) = (g_0 - 1)/(g_0 - 1 + e_0)$ where g_0 is social marginal welfare weight on those out-of-work and e_0 is (minus) the elasticity of the fraction of individuals out-of-work with respect to the retention rate $1 - T'(0)$. Unlike in the standard utilitarian case, in which $g_0 > 1$ and hence $T'(0) > 0$, with free loaders, g_0 is lower and could be lower than one, in which case $T'(0) < 0$, i.e., in-work benefits are optimal.[30]

Application 2: Transfers over the business cycle. Individuals are less likely to be responsible for their unemployment status in a recession than in an expansion, so that the composition of those out of work changes over the business cycle. Formula (5) hints at the fact that benefits might be expanded in bad times. Our online survey presented in Section 5 shows indeed that support for the unemployed depends critically on whether they can or cannot find jobs.

3.4 Tagging and Horizontal Equity Concerns

The standard utilitarian framework leads to the conclusion that if agents can be separated into different groups, based on attributes, so-called "tags," which are correlated with earnings ability, then an optimal tax system should have differentiated taxes for those groups.[31] Mankiw and Weinzierl (2010) explore a tax schedule differentiated by height and use this stark example as a critique of the standard utilitarian framework. In practice, society seems to oppose taxation based on such characteristics, probably because it is deemed unfair to tax differently people with the same ability to pay. These 'horizontal equity' concerns, or the wish to treat 'equals as equals' seem important in practice and a realistic framework for optimal tax policy needs to be able to include them.[32]

It is possible to capture horizontal equity concerns using generalized social welfare weights if we extend our basic framework to allow the weights to be dependent on the *reform considered* (instead of depending solely on the current level of taxes and transfers as we have done so far).

To do this in the simplest way possible, we consider linear taxation designed to raise a given (non-transfer) revenue E as in the Ramsey tax problem. There are two groups which differ according to some observable and perfectly inelastic attribute $m \in \{1, 2\}$ and according to their

[30] Note that if $g_0 < 1 - e_0$, then it is desirable to lower the transfer of non-workers by decreasing $T'(0)$ up to the point where $g_0 - 1 + e_0 > 0$ and the formula holds.

[31] Some attributes can be perfect tags in the sense of being impossible to influence by the agent. An example would be height, which has been shown to be positively correlated with earnings (see Mankiw and Weinzierl, 2010), or gender. Others are potentially elastic to taxes (such as the number of children or marital status).

[32] Kaplow (2001), and Kaplow and Shavell (2001) highlight that Horizontal Equity considerations, in particular as modeled in Auerbach and Hassett (2002) will conflict with the Pareto principle in some cases. Our non-negative social welfare weights guarantee that this cannot occur in our setup. Naturally, the concept of horizontal equity per se remains subject to the valid criticisms raised in Kaplow (2001).

taxable income elasticities (respectively denoted e_1 and e_2). Both groups have measure 1 and the tax rate on group m is denoted by τ_m. The budget constraint, with multiplier p, is:

$$\tau_1 \cdot \int_{i\in 1} z_i + \tau_2 \cdot \int_{i\in 2} z_i \geq E \tag{9}$$

The standard utilitarian approach would naturally lead to different tax rates for each group, following standard Ramsey considerations, such that:

$$\tau_m = \frac{1-\bar{g}_m}{1-\bar{g}_m+e_m} \quad \text{with} \quad \bar{g}_m = \frac{\int_{i\in m} u_{ci}\cdot z_i}{p\cdot \int_{i\in m} z_i},$$

where $p > 0$ is determined so that the tax system raises exactly E to meet the budget constraint (9). Hence, the standard tax system would generate horizontal inequities, i.e., some individuals are taxed more than others based on a tag and conditional on ability to pay measured by income. Without loss of generality, suppose throughout that $e_2 > e_1$ so that group 2 is more elastic and should optimally be taxed less.[33]

Let social marginal welfare weights be concentrated on those suffering from the horizontal inequity. This means that horizontal inequities carry a higher priority than vertical inequities for social justice and that vertical inequity is ignored as long as the tax and transfer system creates horizontal inequity. We also need to impose conditions on weights in situations with no horizontal inequities to ensure that the standard optimum with no tagging remains an equilibrium, i.e., that no small reform can improve welfare. If the small reform creates horizontal inequities, i.e., introduces a small tag, social marginal welfare weights need to depend on the small reform and not only on the current level of taxes and transfers. Weights need to be concentrated on those who suffer from horizontal inequity because of the reform.

Formally, let the weight for person i in group m be $g_i = g(\tau_m, \tau_n, d\tau_m, d\tau_n)$, independent of consumption or work (where n denotes the other group). The weight on a group discriminated against is 1, while the weight on a group that is favored by the tax system is zero. If both groups are fairly treated, weights are uniformly equal to 1. Such weights are designed to avoid horizontal inequities unless they benefit everybody. Absent any horizontal inequity, society has no value for redistribution. More precisely:

i) $g(\tau_m, \tau_n, d\tau_m, d\tau_n) = 1$ and $g(\tau_n, \tau_m, d\tau_n, d\tau_m) = 0$ if $\tau_m > \tau_n$.

ii) $g(\tau, \tau, d\tau_m, d\tau_n) = 1$ and $g(\tau, \tau, d\tau_n, d\tau_m) = 0$ if $\tau_m = \tau_n = \tau$ and $d\tau_m > d\tau_n$.

[33] We assume here that the distribution of income across the two groups is close enough so that any difference between $\bar{g}_1$ and $\bar{g}_2$ is too small to counterbalance the difference in elasticities e_1 and e_2.

iii) $g(\tau, \tau, d\tau_m, d\tau_n) = g(\tau, \tau, d\tau_n, d\tau_m) = 1$ if $\tau_m = \tau_n = \tau$ and $d\tau_m = d\tau_n$.

Regularity assumptions. We assume that there is a uniform tax rate $\tau_1 = \tau_2 = \tau^*$ that can raise E. We assume that the tax functions $\tau_1 \to \tau_1 \cdot \int_{i\in 1} z_i$ and $\tau_2 \to \tau_2 \cdot \int_{i\in 2} z_i$ are single peaked. We also assume that the uniform rate tax function $\tau \to \tau \cdot (\int_{i\in 1} z_i + \int_{i\in 2} z_i)$ is single peaked.

Naturally, the peaks are at $\tau_1 = 1/(1+e_1)$, $\tau_2 = 1/(1+e_2)$, and $\tau = 1/(1+e)$, where e is the elasticity of total income.

Proposition 6 *Let τ^* be the (smallest) uniform rate that raises E: $\tau^*(\int_{i\in 1} z_i + \int_{i\in 2} z_i) = E$.*

i) If $1/(1+e_2) \geq \tau^*$ then the only equilibrium has horizontal equity with $\tau_1 = \tau_2 = \tau^*$.

ii) If $1/(1+e_2) < \tau^*$ then the only equilibrium has horizontal inequity with $\tau_2 = 1/(1+e_2) < \tau^*$ (revenue maximizing rate on group 2) and $\tau_1 < \tau^*$ the smallest tax rate such that $\tau_1 \cdot \int_{i\in 1} z_i + \tau_2 \cdot \int_{i\in 2} z_i = E$. In this case, the equilibrium with horizontal inequity Pareto dominates the uniform tax system $\tau_1 = \tau_2 = \tau^*$.

Therefore, a tax system with horizontal inequity can be an equilibrium only if it helps the group discriminated against, i.e., no reform can help those discriminated against. This happens only when tagging creates a Pareto improvement, which dramatically reduces the scope for using tagging in practice. If the government wants to set τ_2 at a lower level than τ_1, then τ_2 must necessarily be set at the revenue maximizing rate.

This is reminiscent of a Rawlsian setup, in which society only cares about the least well-off. Here, the set of people whom society cares about is endogenous to the tax system. Namely, they are the ones discriminated against because of tagging. In other words, we can rephrase the Rawlsian criterion as follows: "It is permissible to discriminate against a group using taxes and transfers not based on ability to pay only in the case where such discrimination allows to improve the welfare of the group discriminated against."

4 Link with Alternative Justice Principles

In this section, we illustrate how our framework can be connected to justice principles that are not captured by the standard welfarist approach but have been discussed in the normative tax policy literature. We use formula (4) to show how social welfare weights derived from alternative justice principles map into optimal tax formulas.

4.1 Libertarianism, Rawlsianism, and Political Economy

Libertarian case. From the libertarian point of view, any individual is fully entitled to his pre-tax income and society should not be responsible for those with lower earnings. This view could for example be justified in a world where individuals differ solely in their preferences for work but not in their earning ability. In that case, there is no good normative reason to redistribute from consumption lovers to leisure lovers (exactly as there would be no reason to redistribute from apple lovers to orange lovers in an exchange economy where everybody starts with the same endowment). This can be modeled in our framework by assuming that $g_i = g(c_i, z_i) = \tilde{g}(c_i - z_i)$ is increasing in its only argument. Hence, x_i^s and x_i^b are empty. Formula (4) immediately delivers $T'(z_i) \equiv 0$ at the optimum since then $\bar{g}(z) \equiv 1$ and hence $\bar{G}(z) \equiv 1$ when marginal taxes are zero. In the standard framework, the way to obtain a zero tax at the optimum is to either assume that utility is linear or to specify a *convex* transformation of $u(.)$ in the social welfare function which undoes the concavity of $u(.)$.[34]

Rawlsian case. The Rawlsian case is the polar opposite of the Libertarian one. Society cares most about those with the lowest earnings and hence sets the tax rate to maximize their welfare. With a social welfare function, this can be captured by a maximin criterion.[35] In our framework, it can be done instead by assuming that social welfare weights are concentrated on the least advantaged: $g_i = g(u_i - \min_j u_j) = 1(u_i - \min_j u_j = 0)$ so that neither z_i nor c_i (directly) enter the welfare weight and $x_i^s = u_i - \min_j u_j$, while x^b is empty (there could still be heterogeneity in individual characteristics as captured in x_i^u.) If the least advantaged people have zero earnings, independently of taxes, then $\bar{G}(z) = 0$ for all $z > 0$. Formula (4) then implies $T'(z) = 1/[1 + \alpha(z) \cdot e(z)]$ at the optimum. Marginal tax rates are set to maximize tax revenue so as to make the demogrant $-T(0)$ as large as possible.

Political Economy. Political economy considerations can be naturally incorporated. The most popular model for policy decisions among economists is the median-voter model. Consider one specialisation of our general model, with $u_i = u((1-\tau)z_i + \tau \int_i z_i - v(z_i; x_i^u))$. These are single peaked preferences in τ, so that the preferred tax rate of agent i is: $\tau_i = (1 - z_i / \int_i z_i)/(1 - z_i / \int_i z_i + e)$. Hence, the median voter is the voter with median income, denoted by z_m and hence the political equilibrium has: $\tau = \frac{1 - z_m / \int_i z_i}{1 - z_m / \int_i z_i + e}$. Note that $\tau > 0$ when $z_m < \int_i z_i$, which is the standard case with empirical income distributions. This case can be seen as a particular case of endogenous weights where all the weight is concentrated at the median voter.

[34] Weinzierl (2012) incorporates a libertarian element in an optimal tax model by considering a mixed objective.
[35] Atkinson (1975) derives formally the Ralwsian optimal income tax using the maxi-min approach.

4.2 Equality of Opportunity

To capture the concept of equality of opportunity, Roemer (1998) and Roemer et al. (2003) consider models where individuals differ in their ability to earn, but part of the ability is due to family background (which individuals are not responsible for) and to merit (which individuals are responsible for). Conditional on family background, merit could be measured by the percentile of the earnings distribution the individual is in. Society is willing to redistribute across family backgrounds but not across merit (i.e., not across earnings conditional on family background).

Formally, individual utility is $u_i = u(c_i - v(z_i/w_i, B_i))$. w_i is productivity, differences of which across individuals are considered as fair, for example because productivity may have been acquired through hard work in school. $B_i \in \{0, 1\}$ is an individual's background, which can be high ($B_i = 1$) or low ($B_i = 0$). A high family background gives an earnings advantage that is deemed unfair, such that $\partial v(z_i/w_i, 0)/\partial z_i > \partial v(z_i/w_i, 1)/\partial z_i$, for all (z_i, w_i). Let r_i be the percentile of individual i in the earnings distribution conditional on his background, i.e., a measure of the effort or merit of individual i. Hence, conditional on earnings, individuals with a disadvantaged background are more meritorious than advantaged individuals because they are in a higher percentile of the earnings distribution conditional on their background.

Denote by $\bar{c}(r) \equiv (\int_{(i:r_i=r)} c_i)/Prob(i : r_i = r)$ the average consumption of those at rank r. To capture the fact that those from low backgrounds are deemed more deserving, consider welfare weights $g_i = g(c_i; \bar{c}(r_i)) = 1(c_i \leq \bar{c}(r_i))$, with $x_i^s = \bar{c}(r_i)$, $x_i^u = B_i$ and x_i^b empty.

It is assumed that the government cannot observe family background and hence has to base taxes and transfers on earnings z only.[36] Hence, we need to aggregate the marginal welfare weights at each z level. They imply that $\bar{G}(z)$ is the fraction of individuals from disadvantaged background earning at least z relative to the population wide fraction of individuals from disadvantaged background. This is also known as the representation index.

For any background blind tax system, only people from a low background carry a positive weight at any income level, hence the implied social welfare function is:

$$SWF = \int u(z - T(z) - v(z, w))h(z|0)dz$$

where $h(z|0)$ denotes the earnings density of those coming from disadvantaged backgrounds. This corresponds to the social welfare function used by Roemer et al. (2003).[37]

[36] Clearly, a tax conditional on background could erase all earnings differences due to background, similar to the observable luck income case in Section 3.2.

[37] Roemer et al. (2003) does not include disutility of work in its social objective. We include disutility of work

Naturally, we expect $\bar{G}(z)$ to decrease with z as it is harder for those from disadvantaged background to reach upper incomes. If the representation of individuals from disadvantaged backgrounds is zero at the top, the top tax rate should be set to maximize tax revenue. Hence, the theory of Roemer provides a justification for having social welfare weights decreasing with income, which is orthogonal to the utilitarian mechanism of decreasing marginal utility of consumption (as was also the case with our model of deserved and luck income from Section 3.2).

Calibrating the weights to US intergenerational mobility: The recent US intergenerational income mobility statistics produced by Chetty *et al.* (2014) can be used to illustrate this discussion. Suppose we define low background as having parents coming from the bottom 50% of the income distribution. Column (1) in Table 2 displays the fraction of individuals with parents below median income above various percentiles of the income distribution.[38] As individuals with parents below median are by definition half of the population, $\bar{G}(z)$ is simply half of column (1) and is reported in column (2). $\bar{G}(z)$ falls from 100% at percentile 0 (by definition) to 34% at the 99.9th percentile. Hence, in contrast to the standard utilitarian case where $\bar{G}(z)$ converges to zero for large z (with a concave utility function with marginal utility converging to zero), in the equality of opportunity case, $\bar{G}(z)$ converges to a positive value of 1/3 because a substantial fraction of high earners come from disadvantaged backgrounds. $\bar{G}(z)$ appears stable within the 99th percentile as $\bar{G}(z)$ is virtually the same at the 99th percentile and the 99.9th percentile. Hence, under this equality of opportunity criterion, individuals at the 99.9th percentile are deemed no less deserving than individuals at the lower 99th percentile because they are equally likely to come from a disadvantaged background.

This has two important optimal tax consequences for top earners. First, with a Pareto parameter $a = 1.5$ and an elasticity $e = .5$, the optimal top asymptotic tax rate is $\tau = 1/(1+a\cdot e) = 57\%$ in the utilitarian case and $\tau = (1-1/3)/(1-1/3+a\cdot e) = 47\%$ in the meritocratic case, i.e., 10 points lower.[39] Second, a society which values individuals coming from low background would use progressive income taxation but the top tax rate would be stable within the top 1% because the representation of individuals from disadvantaged background is stable within the top 1%. Hence, there should be no additional brackets above the 99th percentile.

To illustrate these properties, Table 2, column (3) presents optimal marginal tax rates at various income levels using formula (4). The weights $\bar{G}(z)$ are from column (2). We calibrate

to be able to apply to our framework and obtain a (constrained) Pareto efficient outcome.

[38] These estimates are based on all US individuals born in 1980-1 with their income measured at age 30-31. In this simulation, we take a short-cut and assume they hold more broadly in the full population.

[39] Naturally, in a less meritocratic society than the United States at present, $\bar{G}(z)$ for large z could possibly be smaller (and the optimal top tax rate correspondingly closer to the optimal utilitarian tax rate).

$\alpha(z)$ using the actual distribution of income based on 2008 income tax return data, the latest year available. We use a constant elasticity $e = 0.5$ which is a mid to upper range estimate based on the literature (Saez, Slemrod, and Giertz, 2012). Because of uncertainty in the level of e, the simulations should be considered as illustrative at best. Column (3) shows that optimal marginal tax rates are U-shaped but about constant above the 99th percentile. For comparison, columns (4) and (5) present the utilitarian weights $\bar{G}(z)$ and optimal marginal tax rates $T'(z)$ assuming a log-utility so that the welfare weight $\bar{g}(z)$ at income level z, is proportional to $1/(z - T(z))$. The utilitarian case delivers optimal tax rates that are about 10 points higher than the equality of opportunity case and significantly more progressive.

4.3 Poverty Alleviation

The poverty rate, defined as the fraction of households below a given disposable income threshold (the poverty threshold) gets substantial attention in the public debate. Hence, it is conceivable that governments aim to either reduce the poverty gap (defined as the amount of money needed to lift all households out of poverty) or reduce the poverty rate (the number of households below the poverty threshold). A few studies have considered government objectives incorporating such poverty concerns. Besley and Coate (1992) and Kanbur, Keen, and Tuomala (1994) show how adopting poverty minimization indexes affects optimal tax analysis. Importantly, their findings imply that the outcomes can be Pareto dominated. In this section, we show how generalized welfare weights can incorporate poverty alleviation considerations in the traditional optimal tax analysis while maintaining the Pareto principle.

Let us denote the poverty threshold by $\bar{c}$. Anybody with disposable income $c < \bar{c}$ is poor. Utility is taken to be a special case of our general formulation: $u_i = u(c_i - v(z_i/w_i))$.

Standard approach and implicit negative weights. The standard objective for poverty reducation is the poverty gap, as studied in Kanbur, Keen, and Tuomala (1994). If the lowest ability agent exerts positive labor supply, the authors find that the bottom marginal tax rate should be negative. It is well-known, however, that in the welfarist case, the optimal tax rate at the bottom is zero (Seade, 1977) and is otherwise non-negative. Indeed, starting from a negative bottom rate, slightly increasing the bottom marginal tax rate would be Pareto improving without violating incentive constraints: It both allows the lowest productivity agent to work less, which is welfare improving, and raises more revenue.[40]

[40] The proof is symmetrical to the proof of the famous zero marginal tax rate top result of Seade and Sadka. A positive top marginal tax rate is Pareto dominated. Proving that the top marginal tax rate is not negative is

The discrepancy between the poverty gap minimization and the welfarist objective arises because the former does not take into account the disutility from work for the lowest productivity agents and finds it profitable to push them to work more.

Generalized Social Welfare Weights. If the demogrant can be made bigger than $\bar{c}$, then the optimum way to fight poverty is to raise enough taxes to set the demogrant equal to $\bar{c}$. Once the poverty threshold has been attained, there is no reason to have differences in social welfare weights and hence the weights would all be equal to a fixed g for those with positive earnings so that $\bar{G}(z) = g$ for $z > 0$. Using formula (4), we would have $T'(z) = (1-g)/(1-g+\alpha(z)\cdot e(z))$, where g is set so that total taxes collected raise enough revenue to fund the demogrant $\bar{c}$. The less trivial case is when even with $g = 0$ (which corresponds to the Rawlsian case), tax revenue cannot fund a demogrant as large as $\bar{c}$. Let us denote by $\bar{z}$ the (endogenous) earnings level such that $\bar{c} = \bar{z} - T(\bar{z})$, i.e., that defines the pre-tax poverty threshold.

Poverty gap alleviation. Suppose the government cares mostly about agents' shortfall in consumption relative to the poverty line. A natural way to capture this is to assume that social welfare weights are concentrated among individuals with disposable income c below the poverty threshold $\bar{c}$. We can therefore specify the generalized welfare weights as follows: $g(c, z; \bar{c}) = 1 > 0$ if $c < \bar{c}$ and $g(c, z; \bar{c}) = 0$ if $c \geq \bar{c}$.[41] In this case, $x_i^u = w_i$, $x_i^s = \bar{c}$, and x^b is empty. We have $\bar{g}(z) = 0$ for $z \geq \bar{z}$ and $\bar{g}(z) = 1/H(\bar{z})$ for $z < \bar{z}$ so that $\int_0^\infty \bar{g}(z)dH(z) = 1$. Hence, we have $\bar{G}(z) = 0$ for $z \geq \bar{z}$ and $\bar{G}(z) = [1 - H(z)/H(\bar{z})]/[1 - H(z)]$ for $z < \bar{z}$. Applying formula (4) yields the following proposition.

Proposition 7 *The optimal tax schedule that minimizes the poverty gap is:*

$$T'(z) = \frac{1}{1+\alpha(z)\cdot e(z)} \quad \text{if} \quad z > \bar{z}$$

$$T'(z) = \frac{(1/H(\bar{z})-1)H(z)}{(1/H(\bar{z})-1)H(z)+\alpha(z)[1-H(z)]\cdot e(z)} \quad \text{if} \quad z \leq \bar{z}$$

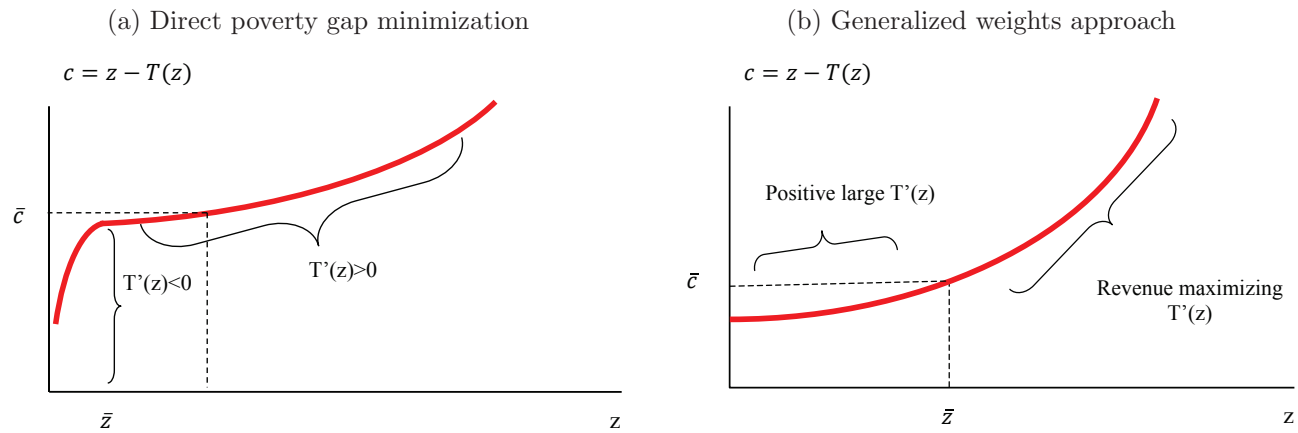
As $(1/H(\bar{z})-1)H(\bar{z}) = 1-H(\bar{z})$, the marginal tax rate is continuous at the poverty threshold $\bar{z}$. The marginal tax rate is Rawlsian above $\bar{z}$ and positive (and typically large) below $\bar{z}$. The shape of optimal tax rates is quite similar to the standard utilitarian case and is illustrated on Figure 1(b) in a (pre-tax income, post-tax income) plane. The case of "poverty rate minimization," in

harder and in fact requires assuming that social marginal welfare weights decrease with income (Seade, 1977).

[41] A less extreme version of this assumption would set $g(c, z; \bar{c}) = \underline{g}$ if $c \geq \bar{c}$ with $0 < \underline{g} < 1$. It is easy to adapt our results to that case.

which the government attempts to minimize the number of people living below the poverty line by concentrating weights on those at the poverty threshold, is treated in Appendix A.5.

Figure 1: Optimal policies for poverty gap minimization



The figure displays the optimal tax schedule for poverty gap alleviation in a (pre-tax income z, post-tax income $c = z - T(z)$) plane. In panel (a), we plot the schedule for the standard approach that consists in directly minimizing the poverty gap. The marginal tax rate is negative below the poverty threshold $\bar{z}$. In panel (b), we plot the schedule derived using generalized welfare weights. The optimal tax schedule has a shape similar to the standard utilitarian case with high marginal tax rates at the bottom.

4.4 Fair Income Taxation

The fair income taxation theory developed by Fleurbaey and Maniquet considers optimal income tax models where individuals differ in skills and in preferences for work.[42] Based on the "Compensation objective" (Fleurbaey, 1994) and the "Responsibility objective", the theory develops social objective criteria that trade-off the "Equal Preferences Transfer Principle" (at the same preferences, redistribution across unequal skills is desirable) and the "Equal Skills Transfer Principle" (at a given level of skill, redistribution across different preferences is not desirable). A trade-off arises because it is impossible to satisfy both principles simultaneously. Intuitively, the government wants to favor the hard working low skilled but cannot tell them apart from the "lazy" high skilled. In this section, we outline how one criterion of fair income tax theory (the $w_{\min}$-equivalent leximin criterion) translates into a profile of social marginal welfare weights. Our outline does not provide complete technical details. We simply reverse engineer

[42] Fleurbaey (2008) and Fleurbaey and Maniquet (2011), chapters 10 and 11 present their fair income tax framework in detail. A number of studies in standard optimal income tax theory has also considered models with heterogeneity in both preferences and skills (see Boadway et al. 2002, Cuff, 2000, Lockwood and Weinzierl, 2012, and the surveys by Kaplow, 2008 and Boadway 2012).

the weights using the optimal fair income tax formula. Fleurbaey and Maniquet (2013) provide (independently) a more rigorous and complete connection between the axioms of fair income tax theory and standard optimal income taxation.[43]

We specialize our general framework to the utility function: $u_i = c_i - v(z_i/w_i, \theta_i)$ where w_i is again the skill of individual i and θ_i captures heterogeneous preferences for work. Hence labor supply is $l_i = z_i/w_i$ and it is assumed that $l \in [0, 1]$ so that $l = 1$ represents full-time work. Again, formula (4) provides the optimal marginal tax rate in this model.

The $w_{\min}$-equivalent leximin criterion proposed by Fleurbaey and Maniquet puts full weight on those with $w = w_{\min}$ who receive the smallest net transfer from the government.

This criterion leads to an optimal tax system with zero marginal tax rates in the earnings range $[0, w_{\min}]$. Therefore, all individuals with earnings $z \in [0, w_{\min}]$ receive the same transfer. The optimal tax system maximizes this transfer and has positive marginal tax rate above $w_{\min}$, with $T'(z) = 1/(1 + \alpha(z) \cdot e(z)) > 0$ for $z > w_{\min}$ (Theorem 11.4 in Fleurbaey and Maniquet, 2011). Using (4), this optimal tax system implies that $\bar{G}(z) = 1$ for $0 \leq z \leq w_{\min}$, i.e., $\int_z^\infty [1 - g(z')]dH(z') = 0$. Differentiating with respect to z, we get $\bar{g}(z) = 1$ for $0 \leq z \leq w_{\min}$. This implies that the average social marginal welfare weight on those earning less than $w_{\min}$ is equal to one. Because the government tries to maximize transfers to those earning less than $w_{\min}$, social marginal welfare weights are zero above $w_{\min}$.[44]

This criterion, and the average weights $g(z)$ implied by it, can be founded on the following underlying generalized social marginal welfare weights. Let $T_{\max} \equiv \max_{(i:w_i=w_{\min})}(z_i - c_i)$. Formally, the weights are functions: $g_i = g(c_i, z_i; w_i, w_{\min}, T_{\max})$ where $x_i^b = w_i$, $x_i^u = \theta_i$, and $x_i^s = (w_{\min}, T_{\max})$, where $w_{\min}$ is an exogenous aggregate characteristic, while $T_{\max}$ is an endogenous aggregate characteristic. Note that, as discussed in the outline of our approach, the characteristics that appear in the utility function but not in the social welfare weights are characteristics that society does not want to redistribute accross. This is the case here for preferences for work θ_i, which are not considered fair to compensate for. This is in contrast to the "Free Loaders" case in section 3.3, where the cost of work was viewed as caused by health differentials or disability, which are considered as fair to compensate for.

More precisely, the weights that rationalize the Fleurbaey-Maniquet tax system are such that: $g(c_i, z_i; w_i, w_{\min}, T_{\max}) = \tilde{g}(z_i - c_i; w_i, w_{\min}, T_{\max})$ with i) $\tilde{g}(z_i - c_i; w_i, w_{\min}, T_{\max}) = 0$ for

[43] Our approach using formula (4) requires estimating weights by income level. It is of course not always straightforward to derive aggregated weights by income level (see Fleurbaey and Maniquet, 2013 for a discussion of this important point).

[44] As social marginal welfare weights $\bar{g}(z)$ average to one, this implies there is a welfare weight mass at $w_{\min}$.

$w_i > w_{\min}$, for any $(z_i - c_i)$ (there is no social welfare weight placed on those with skill above $w_{\min}$ no matter how much they pay in taxes) and ii) $\tilde{g}(.; w_{\min}, w_{\min}, T_{\max})$ is an (endogenous) Dirac distribution concentrated on $z - c = T_{\max}$ (that is, weights are concentrated solely on those with skill $w_{\min}$ who receive the smallest net transfer from the government). This specification forces the government to provide the *same* transfer to all those with skill $w_{\min}$. Otherwise, if an individual with skill $w_{\min}$ received less than others, all the social welfare weight would concentrate on her and the government would want to increase transfers to her. When there are agents with skill level $w_{\min}$ found at every income level below $w_{\min}$, the sole equilibrium is to have equal transfers, i.e., $T'(z) = 0$ in the $[0, w_{\min}]$ earnings range. Weights are zero above earnings $w_{\min}$ as $w_{\min}$-skilled individuals can at most earn $w_{\min}$, even when working full time.

5 Empirical Testing using Survey Data

The next step in this research agenda is to provide empirical foundations for our theory. The basic tool we use is a series of online survey questions destined to elicit people's preferences for redistribution and their concepts of fairness. The questions are clustered in two main groups. The first set serves to find out what notions of fairness people use to judge tax and transfer systems. We focus on the themes addressed in this paper, such as taxes paid matter (keeping disposable income constant), whether the wage rate and hours of work matter (keeping earned income constant), or whether transfer recipients are perceived to be more or less deserving based on whether they can work or not. The second set has a more quantitative ambition. As described in Section 3, it aims at estimating whether and how social marginal welfare weights depend both on disposable income c and taxes paid T.

Our survey was conducted in December 2012 on Amazon's Mechanical Turk service, using a sample of slightly more than 1100 respondents.[45] The complete details of the survey are presented in appendix. The survey asks subjects to tell which of two families (or individuals) are most deserving of a tax break (or a benefit increase). The families (or individuals) differ in earnings, taxes paid, or other attributes.

5.1 Qualitative Social Preferences

Table 3 reports preferences for giving a tax break and or a benefit increase across individuals in various scenarios.

[45] The full survey is available online at `https://hbs.qualtrics.com/SE/?SID=SV_9mHljmuwqStHDOl`

Marginal utility of income. Panel A considers two individuals with the same earnings, same taxes, and same disposable income but who differ in their marginal utility of income. One person is described as "She greatly enjoys spending money, going out to expensive restaurants, or traveling to fancy destinations. She always feels that she has too little money to spend." while the other person is described as "She is a very frugal person who feels that her current income is sufficient to satisfy her needs." Under standard utilitarianism, the consumption loving person should be seen as more deserving of a tax break than the frugal person. In contrast, 74.4% of people report that consumption loving is irrelevant suggesting that marginal utilities driven by individual taste should not be relevant for tax policy as long as disposable income is held constant. This fits with the view described in this paper that, in contrast to welfarism, actual social welfare weights have little to do with tastes for enjoying consumption. Furthermore, in sharp contrast to utilitarianism, 21.5% think the frugal person is most deserving and only 4.4% of people report that the consumption loving person is the most deserving of a tax break. This result is probably due to the fact that, in moral terms, "frugality" is perceived as a virtue while "spending" is perceived as an indulgence.

Hard worker vs. leisure lover. Panel B considers two individuals with the same earnings, same taxes, and same disposable income but different wage rates and hence different work hours: one person works 60 hours a week at $10 per hour while the other works only 20 hours a week at $30 per hour. 54.4% of respondents think hours of work is irrelevant. This suggests again that for a majority (albeit a small one), hours of work and wage rates are irrelevant for tax policy as long as earnings are the same. A fairly large group of 42.7% of subjects think the hardworking low wage person is more deserving of a tax break while only 2.9% think the part-time worker is most deserving. This provides support to the fair income tax social criteria of Fleurbaey and Maniquet discussed in Section 4.4. Long hours of work do seem to make a person more deserving than short hours of work, conditional on having the same total earnings.

Transfer recipients and free loaders. Panel C considers transfer recipients receiving the same benefit levels. Subjects are asked to rank 4 individuals in terms of deservedness of extra benefits: (1) a disabled person unable to work, (2) an unemployed person actively looking for work, (3) an unemployed person not looking for work, (4) a welfare recipient not looking for work. Subjects rank deservedness according to the order just listed. In particular, subjects find the disabled person unable to work and the unemployed person looking for work much more deserving than the able-bodied unemployed or welfare recipient not looking for work. This provides very strong support to the "free loaders" theory laid out in Section 3.3 that ability

and willingness to work are the key determinants of deservedness of transfer recipients. These results are consistent with a broad body of work discussed above.

Disposable income vs. taxes paid. In the spirit of our analysis of Section 3 with fixed incomes, we analyze whether revealed social marginal welfare weights depend on disposable income and/or taxes paid. Table 4 presents non-parametric evidence showing that both disposable income and taxes paid matter and hence that subjects are neither pure utilitarians (for whom only disposable income matters) nor pure libertarians (for whom only taxed paid matter).

Panel A in Table 4 considers two families A and B with similar disposable income but dissimilar pre-tax income (and hence, different taxes paid). Family B has lower taxes and pre-tax incomes than family A. We keep family B constant and vary family A's taxes and disposable income. Overall, subjects overwhelmingly find family A more deserving than family B. To put it simply, most people find that a family earning $50,000 and paying $15,000 in taxes is more deserving of a tax break than a family earnings $40,000 and paying $5,000 in taxes. This implies that disposable income is not a sufficient statistics to determine deservedness, and that taxes paid enter deservedness positively. This contradicts the basic utilitarian model of Section 3.1.

One small caveat in this interpretation is that if respondents consider consumption and labor to be complementary in utility, they might be choosing to compensate people who earn more income through higher consumption. However, as shown by Chetty (2006), labor supply fluctuations are not very correlated with consumption changes, so that consumption and labor cannot be complementary enough to explain our results.

Panel B in Table 4 considers two families A and B with similar taxes paid but dissimilar pre-tax income (and hence dissimilar disposable income as well). Family B has lower pre-tax and disposable income than family A. We again keep family B constant and vary family A taxes and disposable income. Subjects overwhelmingly find family B more deserving than family A. To put it simply, most people find that a family earning $40,000 and paying $10,000 in taxes is more deserving of a tax break than a family earnings $50,000 and paying $10,000 in taxes. This implies that taxes paid is not a sufficient statistics to determine deservedness and that disposable income affects deservedness negatively. This contradicts the basic libertarian model.

Therefore, Table 4 provides compelling non-parametric evidence that both taxes and disposable income matter for social marginal welfare weights as we posited in Section 3.

5.2 Quantifying Social Preferences

Table 5 provides a first attempt at estimating the weights placed by social preferences on both disposable income and taxes paid. Recall the simple linear form discussed above, $\tilde{g}(c,T) = \tilde{g}(c - \alpha T)$, for which the optimal marginal tax rate with no behavioral effects is constant at all income levels and equal to $T' = 1/(1+\alpha)$. To calibrate α, we created 35 fictitious families, each characterized by a level of taxes and a level of net income.[46] Respondents were sequentially shown five pairs, randomly drawn from the 35 fictitious families and asked which family is the most deserving of a \$1,000 tax break. This menu of choices allows us in principle to recover the social preferences $\tilde{g}(c,T)$ of each subject respondent.

Define a binary variable S_{ijt} which is equal to 1 if fictitious family i was selected during random display t for respondent j, and 0 otherwise. The regression studied is:

$$S_{ijt} = \beta_0 + \beta_T dT_{ijt} + \beta_c dc_{ijt},$$

where dT_{ijt} is the difference in tax levels and dc_{ijt} is the difference in net income levels between the two fictitious families in the pair shown during display t to respondent j. Under our assumption on the weights, $dc/dT = \alpha$ represents the slope of the (linear) social indifference curves in the (T,c) space. Families (that is, combinations of c and T) on higher indifference curves have a higher probability of being selected by social preferences. Hence, there is a mapping from the level of social utility derived from a pair (T,c) and the probability of being selected as most deserving in our survey design. The constant slope of social preferences, α, can then be inferred from the ratio $\frac{dc}{dT}|_{S=\text{constant}} = -\frac{\beta_T}{\beta_c}$. Table 5 shows the implied α and the optimal marginal tax rates in four subsamples.[47] The implied α is between 0.37 and 0.65, so that the implicit optimal marginal tax rates are relatively high, ranging from 61% to 74%. In part, this reflects our implicit assumption of no behavioral effects, which would otherwise tend to reduce the optimal tax rates at any given level of redistributive preferences. Interestingly, the implied marginal tax rates decrease when higher income fictitious families are not considered. Columns 5 and 6 highlight an interesting heterogeneity between respondents who classify themselves as "liberal" or "very liberal" (in column 5), and those who classify themselves as "conservative" or "very conservative" (in column 6). Liberals' revealed preferred marginal tax rate is 85%, while

[46] Annual incomes could take one of 7 values \$10K, \$25K, \$50K, \$100K, \$200K, \$500K, \$1 million, and taxes paid (relative to income) could take one of 5 values, 5%, 10%, 20%, 30%, and 50%.

[47] First, using the full sample and then dropping higher income groups (\$1 million and above and \$500K and above respectively) or the lowest income group (\$10K).

that of conservatives is much lower at 57%.

This simple exercise confirms the results from Table 4 that both net income and the tax burden matter significantly for social preferences and that it is possible to determine the relative weight placed on each. More complex and detailed survey work in this spirit could help calibrate the weights more precisely.

6 Conclusion

This paper has shown that the concept of *generalized marginal social welfare weights* is a fruitful way to extend the standard welfarist theory of optimal taxation. The use of suitable generalized social welfare weights can help resolve many of the puzzles of the traditional welfarist approach and account for existing tax policy debates and structures while retaining (local) Pareto constrained efficiency. Our theory brings back social preferences as a critical element for optimal tax theory analysis. Naturally, this flexibility of generalized social weights begs the question of what social welfare weights ought to be and how they are formed.

Generalized welfare weights can be derived from social justice principles, leading to a normative theory of taxation. The most famous example is the Rawlsian theory where the generalized social marginal welfare weights are concentrated solely on the most disadvantaged members of society. As we have discussed, binary weights (equal to one for those deserving more support and zero otherwise) have normative appeal and can be used in a broad range of cases. The Rawlsian case can also be extended to a discrete number of groups, ranked according to deservedness, such that society has redistributive preferences across groups but libertarian preferences within groups. Naturally, who is deserving might itself be endogenous to the tax system. Such weights can also prioritize justice principles in a lexicographic form.

First, injustices created by tax policy (such as violations of horizontal equity) may have the highest priority. In that case, those deserving of support are those discriminated against whenever horizontal inequities arise. This implies that horizontal inequities can only arise if they help the group discriminated against, dramatically lowering the scope for such policies (such as tagging) that are recommended by the standard welfarist approach and that are typically not observed in the real world.

Second, deserving individuals will be those who face difficult economic situations through no fault of their own. This captures the principle of compensation. Health shocks come to mind, explaining why virtually all advanced countries adopt generous public health insurance

that effectively compensate individuals for the bad luck of facing high health expenses. Once disparities in health care costs have been compensated by public health insurance provision, this element naturally drops out of social welfare weights. Family background is obviously another element that affects outcomes and that individuals do not choose. This explains why equality of opportunity has wide normative appeal both among liberals and conservatives. Policies aiming directly to curb such inequities such as public education or inheritance taxation can therefore be justified on such grounds. Naturally, public education or inheritance taxation cannot fully erase inequalities due to family background. This leaves a role for taxes and transfers based on income that aim at correcting for remaining inequities in opportunity as in the theory of Roemer et al. (1993) which can be implemented using intergenerational mobility statistics.

Third, even conditional on background, there remains substantial inequality in incomes. Part of this inequality is due to choices (preferences for leisure vs. consumption) but part is due to luck (ability and temperament are often not based on choice). Naturally, there is a debate on the relative importance of choices vs. luck, which impacts the resulting social welfare weights. As in the fair income tax theory, the generalized social welfare weights have the advantage of highlighting which differences society considers unfair (for example, due to intrinsic skill differences) and which it considers fair (for example, due to different preferences for work).

Finally, there might be scope for redistribution based on more standard utilitarian principles, i.e., the fact that an additional dollar of consumption matters more for lower income individuals than for higher income individuals. In the public debate, this principle seems relevant at the low income end to justify the use of anti-poverty programs but is not widely invoked to justify progressive taxation at the upper end (Mankiw 2010, 2013).

Social preferences are indeed shaped by beliefs about what drives disparities in individual economic outcomes (effort, luck, background, etc.) as in the model of Piketty (1995). As we have shown, online surveys can be used to estimate empirically actual social preferences, leading to a positive theory of taxation. More ambitiously, economists can also cast light on those mechanisms and hence enlighten public perceptions so as to move the debate up to the higher level of normative principles.

Table 1: Generalized Social Marginal Welfare Weights

	Actual practice	Standard Welfarist Criterion	Generalized Social Marginal Welfare Weights (denoted by g)
	(1)	(2)	(3)
Pareto efficiency	Desirable	Yes	Yes (local) if g are not negative
Optimal taxes with fixed incomes	Non-degenerate	Degenerate (full redistribution desirable)	Non-degenerate if g depend directly on taxes paid (in addition to consumption)
Luck income vs. deserved income	Important	Cannot be distinguished	Can be distinguished if g depends on luck vs. deserved income
Free loaders	Important	Cannot be captured	Can be captured if g depends on hypothetical behavior (work or not absent transfers)
Tagging	Used minimally	Highly desirable	Can be made undesirable if g depends on horizontal inequities (g needs to depend on small tax reform)

Note: This table contrasts actual practice (column 1), the standard welfarist approach (column 2), and our generalized social marginal welfare weights approach (column 3) in various situations listed on the left-hand-side of the table. In each situation, column 3 indicates what property of social marginal welfare weights (denoted by g) is required to make this approach fit with actual tax policy practice.

Table 2: Equality of Opportunity vs. Utilitarian Optimal Tax Rates

	Equality of Opportunity			Utilitarian (log-utility)	
	Fraction from low background (=parents below median) above each percentile	Implied social welfare weight G(z) above each percentile	Implied optimal marginal tax rate at each percentile	Utilitarian social welfare weight G(z) above each percentile	Utilitarian optimal marginal tax rate at each percentile
	(1)	(2)	(3)	(4)	(5)
Income percentile					
z= 25th percentile	44.3%	0.886	53%	0.793	67%
z= 50th percentile	37.3%	0.746	45%	0.574	58%
z= 75th percentile	30.3%	0.606	40%	0.385	51%
z= 90th percentile	23.6%	0.472	34%	0.255	42%
z= 99th percentile	17.0%	0.340	46%	0.077	54%
z= 99.9th percentile	16.5%	0.330	47%	0.016	56%

Notes: This table compares optimal marginal tax rates at various percentiles of the distribution (listed by row) using an equality of opportunity criterion (in column (3)) and a standard utilitarian criterion (in column (5)). Both columns use the optimal tax formula $T'(z)=[1-G(z)]/[1-G(z)+\alpha(z)*e]$ discussed in the text where G(z) is the average social marginal welfare weight above income level z, $\alpha(z)=(zh(z))/(1-H(z))$ is the local Pareto parameter (with h(z) the density of income at z, and H(z) the cumulative distribution), and e the elasticity of reported income with respect to 1-T'(z). We assume e=0.5. We calibrate $\alpha(z)$ using the actual distribution of income based on 2008 income tax return data. For the equality of opportunity criterion, G(z) is the representation index of individuals with income above z who come from a disadvantaged background (defined as having a parent with income below the median). This representation index is estimated using the national intergenerational mobility statistics of Chetty et al. (2013) based on all US individuals born in 1980-1 with their income measured at age 30-31. For the utilitarian criterion, we assume a log-utility so that the social welfare weight g(z) at income level z is proportional to 1/(z-T(z)).

Table 3: Revealed Social Preferences

	(1)	(2)	(3)	(4)
A. Consumption lover vs. Frugal				
	Consumption lover > Frugal	Consumption lover = Frugal	Consumption lover < Frugal	
# obs. = 1,125	4.1%	74.4%	21.5%	
	(0.6%)	(1.3%)	(1.2%)	
B. Hardworking vs. leisure lover				
	Hardworking > Leisure lover	Hardworking = Leisure lover	Hardworking < Leisure lover	
# obs. = 1,121	42.7%	54.4%	2.9%	
	(1.5%)	(1.5%)	(0.5%)	
C. Transfer Recipients and free loaders				
# obs. = 1,098	Disabled person unable to work	Unemployed looking for work	Unemployed not looking for work	Welfare recipient not looking for work
Average rank (1-4) assigned	1.4	1.6	3.0	3.5
	(0.018)	(0.02)	(0.023)	(0.025)
% assigned first rank	57.5%	37.3%	2.7%	2.5%
	(1.3%)	(1.3%)	(0.4%)	(0.4%)
% assigned last rank	2.3%	2.9%	25.0%	70.8%
	(0.4%)	(0.4%)	(1.1%)	(1.2%)

Notes: This table reports preferences for giving a tax break and or a benefit increase across individuals in various scenarios. Panel A considers two individuals with the same earnings, same taxes, and same disposable income but high marginal utility of income (consumption lover) vs. low marginal utility of income (frugal). In contrast to utilitarianism, 74.4% of people report that consumption loving is irrelevant and 21.5% think the frugal person is most deserving. Panel B considers two individuals with the same earnings, same taxes, and same disposable income but different wage rates and hence different work hours. 54.4% think hours of work is irrelevant and 42.7% think the hardworking low wage person is more deserving. Panel C considers transfer recipients receiving the same benefit levels. Subjects find the disabled person unable to work and the unemployed person looking for work much more deserving than the abled bodied unemployed or welfare recipient not looking for work. For all statistics, standard errors are reported in parentheses below each estimate.

Table 4: Utilitarian vs. Libertarian Preferences

	(1)	(2)	(3)
A. Utilitarian Test			
	Family B: z=$40,000, T=$5,000, c=$35,000		
Most deserving family	Family A: z=$50,000, T=$14,000, c=$36,000	Family A: z=$50,000, T=$15,000, c=$35,000	Family A: z=$50,000, T=$16,000, c=$34,000
A>B	48.8%	54.8%	65.2%
	(1.5%)	(1.5%)	(1.4%)
A=B	38.8%	37.3%	28.0%
	(1.4%)	(1.4%)	(1.3%)
A<B	12.4%	7.9%	6.8%
	(1.0%)	(0.8%)	(0.7%)
B. Libertarian Test			
	Family B: z=$40,000, T=$10,000, c=$30,000		
Most deserving family	Family A: z=$50,000, T=$11,000, c=$39,000	Family A: z=$50,000, T=$10,000, c=$40,000	Family A: z=$50,000, T=$9,000, c=$41,000
A>B	7.7%	3.6%	3.1%
	(0.8%)	(0.6%)	(0.5%)
A=B	29.1%	40.0%	23.7%
	(1.3%)	(1.5%)	(1.3%)
A<B	63.2%	56.4%	73.2%
	(1.4%)	(1.5%)	(1.3%)

Notes: Sample size 1,111 subjects who finished the survey. Subjects were asked which of Family A vs. Family B was most deserving of a $1,000 tax break in 6 scenarios with different configurations for pre-tax income z, taxes paid T, and disposable income c=z-T. The table reports the fraction of subjects reporting that family A is more deserving (A>B), families A and B are equally deserving (A=B), family B is more deserving (A<B). Standard errors are in parentheses.

Table 5: Calibrating Social Welfare Weights

	Probability of being deemed more deserving in pairwise comparison					
Sample	Full	Excludes cases with income of $1m	Excludes cases with income of $500K+	Excludes cases with income $500K+ and $10K or less	Liberal subjects only	Conservative subjects only
	(1)	(2)	(3)	(4)	(5)	(6)
d(Tax)	0.0017***	0.0052***	0.016***	0.015***	0.00082***	0.0032***
	(0.0003)	(0.0019)	(0.0019)	(0.0022)	(0.00046)	(0.00068)
d(Net Income)	-0.0046***	-0.0091***	-0.024***	-0.024***	-0.0048***	-0.0042***
	(0.00012)	(0.00028)	(0.00078)	(0.00094)	(0.00018)	(0.00027)
Number of observations	11,450	8,368	5,816	3,702	5,250	2,540
Implied α	0.37	0.58	0.65	0.64	0.17	0.77
	(0.06)	(0.06)	(0.07)	(0.09)	(0.12)	(0.16)
Implied marginal tax rate	73%	63%	61%	61%	85%	57%

Notes: Survey respondents were shown 5 randomly selected pairs of fictitious families, each characterized by levels of net income and tax, for a total of 11,450 observations, and asked to select the family most deserving of a $1,000 tax break. Gross income was randomly drawn from {10K, 25K, 50K, 100K, 200K, 500K, 1 mil} and tax rates from {5%, 10%, 20%, 30%, 50%}. The coefficients are from an OLS regression of a binary variable equal to 1 if the fictitious family was selected, on the difference in tax levels and net income levels between the two families of the pair. Column (1) uses the full sample. Column (2) excludes fictitious families with income of 1 mil. Column (3) excludes families with income of 500K or more. Column (4) further excludes in addition families with income below 10K. Column (5) shows the results for all families but only for respondents who classify themselves as "liberal" or "very liberal", while Colum (6) shows the results for respondents who classify themselves as "conservative" or "very conservative". The implied α is obtained as (the negative of) the ratio of the coefficient on d(Tax) over the one on d(Net income). Bootstrap standard errors in parentheses. The optimal implied constant marginal tax rate (MTR) under the assumption of no behavioral effects is, as in the text, MTR = $1/(1+\alpha)$. The implied MTRs are high, between 61% and 74%, possibly due to the assumption of no behavioral effects. In addition, the implied MTR declines when respondents are not asked to consider higher income fictitious families. Respondents who consider themselves Liberals prefer higher marginal tax rates than those who consider themselves Conservatives.

References

Ackert, Lucy, Jorge Martinez-Vazquez, and Mark Rider. 2007. "Social Preferences and Tax Policy Design: Some Experimental Evidence," *Economic Inquiry* 45(3), 487-501.

Alesina, Alberto and George-Marios Angeletos. 2005. "Fairness and Redistribution," *American Economic Review* 95(3), 960-980.

Alesina, Alberto and Paola Giuliano. 2011. "Preferences for Redistribution," in A. Bisin and J. Benhabib (eds.), *Handbook of Social Economics,* Amsterdam: North Holland, Chapter 4, 93–132.

Atkinson, Anthony B. 1975. "La 'Maxi-Min' et l'imposition optimale des revenus," *Cahiers du Séminaire d'Econométrie* 16, 73–86.

Auerbach, Alan and Kevin Hassett. 2002. "A New Measure of Horizontal Equity," *American Economic Review*, 92(4), 1116-1125.

Battaglini, Marco and Steven Coate. 2008. "Pareto Efficient Income Taxation with Stochastic Abilities," *Journal of Public Economics*, 92(3-4), 844–868.

Besley, Timothy and Steven Coate. 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs," *American Economic Review*, 82(1), 249–261.

Besley, Timothy and Stephen Coate. 1992b. "Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination," *Journal of Public Economics* 48(2), 165–183.

Besley, Timothy, Steven Coate, and Ravi Kanbur. 1988. "Food Subsidies and Poverty Alleviation", *Economic Journal,* 92, 701–719.

Boadway, Robin. 2012. *From Optimal Tax Theory to Tax Policy,* Cambridge: MIT Press.

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), 475–798.

Bourguignon, François and Amedeo Spadaro. 2012. "Tax-benefit Revealed Social Preferences," *Journal of Economic Inequality* 10(1), 75-108.

Chetty, Raj. 2006. "A New Method of Estimating Risk Aversion," *American Economic Review* 96(5), 1821-1834.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods."*Annual Review of Economics,* 1, 451–488.

Chetty, Raj, Nathan Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics* 129(4).

Christiansen, Vidar and Eilev Jansen. 1978. "Implicit Social Preferences in the Norwegian System of Indirect Taxation," *Journal of Public Economics* 10(2), 217–245.

Cowell, Frank and Erik Schokkaert. 2001. "Risk perceptions and distributional judgments," *European Economic Review* 45 (4-6), 941-952.

Cuff, Katherine. 2000. "Optimality of Workfare with Heterogeneous Preferences," *Canadian Journal of Economics* 33, 149–174.

Devooght, Kurt and Erik Schokkaert. 2003. "Responsibility-sensitive Fair Compensation in Different Cultures," *Social Choice and Welfare* 21, 207-242

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

Edgeworth, Francis Y. 1897. "The Pure Theory of Taxation," *Economic Journal* 7, 46-70, 226-238, and 550-571.

Ellwood, David. 1988. *Poor Support: Poverty and the American Family*, New York: Basic Books.

Ellwood, David and Mary J. Bane. 1996. *Welfare Realities: from Rhetoric to Reform*, Cambridge: Harvard University Press.

Engelmann, Dirk and Martin Strobel. 2004. "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments," *American Economic Review* 94(4), 857–69.

Farhi, Emmanuel, and Ivan Werning. 2013. "Estate Taxation with Altruism Heterogeneity," *American Economic Review* 103(3), 489–495.

Feldstein, Martin. 2012. "The Mirrlees Review," *Journal of Economic Literature* 50(3), 781–790.

Fleurbaey, Marc. 1994. "On Fair Compensation", *Theory and Decision* 36, 277–307.

Fleurbaey, Marc. 2008. *Fairness, Responsability and Welfare*, Oxford: Oxford University Press.

Fleurbaey, Marc and François Maniquet. 2011. *A Theory of Fairness and Social Welfare*, Cambridge: Cambridge University Press.

Fleurbaey, Marc and François Maniquet. 2013. "The Contribution of Fairness Principles to Optimal Taxation Theory," Working Paper.

Fong, Christina. 2001. "Social Preferences, Self-interest, and the Demand for Redistribution," *Journal of Public Economics* 82(2), 225–246.

Frohlich, Norman and Joe A. Oppenheimer. 1992. *Choosing Justice: An Experimental Approach to Ethical Theory*, Berkeley University of California Press.

Gaertner, Wulf and Erik Schokkaert. 2012. *Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice,* Cambridge: Cambridge University Press.

Golosov, Mikhail, Aleh Tsyvinski and Nicolas Werquin. 2013. "Dynamic Tax Reforms," Yale University Working Paper.

Golosov, Mikhail and Aleh Tsyvinski. 2014. "Policy Implications of Dynamic Public Finance," Annual Review, 7. *In Preparation.*

Guesnerie, Roger. 1995. *A Contribution to the Pure Theory of Taxation*, Cambridge University Press: Cambridge.

Hendren, Nathan. 2013. "The Policy Elasticity," NBER Working Paper No. 19177.

Jeene, Marjolein, Wim van Oorschot, and Wilfred Uunk. 2011. "Popular Criteria for the Welfare Deservingness of Disability Pensioners: The Influence of Structural and Cultural Factors," *Journal of Social Indicators Research*, 1-15.

Kanbur, Ravi, Michael Keen, and Matti Tuomala. 1994. "Optimal Nonlinear Income Taxation for the Alleviation of Income-Poverty," *European Economic Review* 38(8), 1613-1632.

Kaplow, Louis. 2001. "Horizontal Equity: New Measures, Unclear Principles (Commentary)," *Inequality and Tax Policy,* Hassett and Hubbard, eds., American Enterprise Institute, 75–97.

Kaplow, Louis. 2008. *The Theory of Taxation and Public Economics*, Princeton University Press: Princeton.

Kaplow, Louis, and Steven Shavell. 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle," *Journal of Political Economy* 109(2), 281–86.

Kleven, Henrik and Claus Kreiner. 2006. "The Marginal Cost of Public Funds: Hours of Work Versus Labor Force Participation," *Journal of Public Economics* 90, 1955–1973.

Kuziemko, Ilyana, Michael Norton, Emmanuel Saez, and Stefanie Stantcheva. 2013. "How Elastic are Preferences for Redistribution? Evidence from Randomized Survey Experiments", NBER Working paper No. 18865.

Larsen, Christian Albrekt. 2008. "The Institutional Logic of Welfare Attitudes: How Welfare Regimes Influence Public Support," *Comparative Political Studies* 41(2), 145-168.

Lehmann, Etienne and Laurence Jacquet. 2013. "Optimal Nonlinear Income Tax when Skills and Labor Supply Elasticities are Heterogeneous," Working Paper.

Lockwood, Benjamin and Matthew Weinzierl. 2012. "De Gustibus non est Taxandum: Theory and Evidence on Preference Heterogeneity and Redistribution," NBER Working Paper No. 17784.

Mankiw, Gregory. 2010. "Spreading the Wealth Around: Reflections Inspired by Joe the Plumber," *Eastern Economic Journal* 36(3), 285–98.

Mankiw, Gregory. 2013. "Defending the One Percent," *Journal of Economic Perspectives*, 27(3), 21–34.

Mankiw, Gregory and Matthew Weinzierl. 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2(1), 155-76.

Mankiw, Gregory, Matthew Weinzierl, and Danny Yagan. 2009. "Optimal Taxation in Theory and Practice,"*Journal of Economic Perspectives* 23, 147–174.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*, 38: 175-208.

Nozick, Robert. 1974. *Anarchy, State, and Utopia.* Basic Books.

Piketty, Thomas. 1995. "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 110(3), 551-584.

Piketty, Thomas and Emmanuel Saez. 2013. "Optimal Labor Income Taxation," *Handbook of Public Economics,* Volume 5, (Amsterdam: North Holland).

Piketty, Thomas and Emmanuel Saez. 2013b. "A Theory of Optimal Inheritance Taxation," *Econometrica* 81(5), 1851–1886.

Roemer, John. 1998. *Equality of Opportunity*, Cambridge: Harvard University Press.

Roemer, John et al., 2003. "To What Extent Do Fiscal Systems Equalize Opportunities for Income Acquisition Among Citizens?", *Journal of Public Economics*, 87, 539-565.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies* 68, 205-229.

Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive Versus Extensive Labour Supply Responses." *Quarterly Journal of Economics*, 117(2): 1039-73.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature* 50(1), 3–50.

Seade, Jesus. 1977. "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7(2), 203–235.

Weinzierl, Matthew C. 2012. "The Promise of Positive Optimal Taxation," NBER Working Paper No. 18599.

Werning, Ivan. 2007. "Pareto Efficient Taxation," MIT working paper

Will, Jeffrey. 1993. "The Dimensions of Poverty: Public Perceptions of the Deserving Poor," *Social Science Research* 22, 312–332.

Yaari, Menahem E. and Maya Bar-Hillel. 1984. "On Dividing Justly," *Social Choice and Welfare* 1(1), 1–24.

Zoutman, Floris, Bas Jacobs, and Egbert Jongen. 2012. "Revealed Social Preferences of Dutch Political Parties", Tinbergen Institution Discussion Paper.

Online Appendix

A.1 Derivation of the Optimal Tax Formulas using Weights

We show how to derive the optimal nonlinear tax formula (4) and the optimal linear tax formula (5) using the generalized welfare weight approach. In each case, we consider a budget neutral tax reform. At the optimum, the net welfare effect has to be zero.

Optimal non-linear tax (Result 1). Consider a small reform $dT(z)$ in which the marginal tax rate is increased by $d\tau$ in a small band from z to $z + dz$, but left unchanged anywhere else. The reform mechanically collects extra taxes $dzd\tau$ from each taxpayer above z. As there are $1 - H(z)$ individuals above z, $dzd\tau[1 - H(z)]$ is collected. With no income effects on labor supply, there is no behavioral response above the small band.

Those in the income range from z to $z + dz$ have a behavioral response to the higher marginal tax rate. A taxpayer in the small band reduces her income by $\delta z = -ezd\tau/(1-T'(z))$ where e is the elasticity of earnings z with respect to the net-of-tax rate $1 - T'$. As there are $h(z)dz$ taxpayers in the band, those behavioral responses lead to a tax loss equal to $-dzd\tau \cdot h(z)e(z)zT'(z)/(1 - T'(z))$ with $e(z)$ the average elasticity in the small band. Hence, the net revenue collected by the reform is

$$dR = dzd\tau \cdot \left[1 - H(z) - h(z) \cdot e(z) \cdot z \cdot \frac{T'(z)}{1 - T'(z)}\right].$$

This revenue is rebated lumpsum so that the reform is budget neutral.[48]

What is the effect of the reform on welfare using the generalized welfare weights g_i? The welfare effect is $-\int_i g_i dT(z_i)$ with $dT(z_i) = -dR$ for $z_i \leq z$ and $dT(z_i) = d\tau dz - dR$ for $z_i > z$. Hence, the net effect on welfare is $dR \cdot \int_i g_i - d\tau dz \int_{\{i:z_i \geq z\}} g_i$. At the optimum, the net welfare effect is zero. Using the expression for dR above and the fact that $(1 - H(z))\bar{G}(z) = \int_{\{i:z_i \geq z\}} g_i / \int_i g_i$, the net welfare effect can be rewritten as

$$dzd\tau \cdot \int_i g_i \cdot \left[1 - H(z) - h(z) \cdot e(z) \cdot z \cdot \frac{T'(z)}{1 - T'(z)}\right] - dzd\tau \cdot \int_i g_i \cdot (1 - H(z)) \cdot \bar{G}(z) = 0.$$

Dividing by $dzd\tau \cdot \int_i g_i$ and re-arranging, we get

$$\frac{T'(z)}{1 - T'(z)} = \frac{1}{e(z)} \cdot \frac{1 - H(z)}{z \cdot h(z)} \cdot (1 - \bar{G}(z)).$$

Using the local Pareto parameter $\alpha(z) = zh(z)/(1 - H(z))$, we obtain formula (4) in Result 1.

[48] With no income effects, this lumpsum rebate has no labor supply effect on earnings.

Optimal linear tax (Result 2). Consider a small reform $d\tau$. This increases mechanically tax revenue by $d\tau \cdot \int_i z_i$. By definition of the aggregate elasticity e of $\int_i z_i$ with respect to $1-\tau$, this reduces tax revenue through behavioral responses by $-e \cdot \frac{\tau}{1-\tau} \cdot d\tau \cdot \int_i z_i$. Hence, the net effect on revenue is $dR = [1 - e \cdot \frac{\tau}{1-\tau}] \cdot d\tau \cdot \int_i z_i$. This revenue is rebated lumpsum to individuals so that the reform is budget neutral.[49]

What is the effect of the reform on welfare using the generalized welfare weights g_i? The welfare effect $-\int_i g_i dT(z_i)$ with $dT(z_i) = -dR + z_i \cdot d\tau$. Hence, the welfare effect is $dR \cdot \int_i g_i - d\tau \int_i z_i \cdot g_i$. At the optimum, this is zero. Using the expression for dR above, this implies:

$$\left[1 - e \cdot \frac{\tau}{1-\tau}\right] \cdot d\tau \cdot \int_i z_i \cdot \int_i g_i = d\tau \cdot \int_i z_i \cdot g_i$$

or equivalently

$$1 - e \cdot \frac{\tau}{1-\tau} = \bar{g} \quad \text{with} \quad \bar{g} = \frac{\int_i z_i \cdot g_i}{\int_i z_i \cdot \int_i g_i}$$

which can easily be re-expressed as the optimal formula (5) in Result 2.

A.2 Taxation with fixed incomes

Proof of Proposition 4: To see this, suppose by contradiction that $\tilde{g}(z_1 - T(z_1), T(z_1)) > \tilde{g}(z_2 - T(z_2), T(z_2))$. Then transferring a dollar from those earning z_2 toward those earning z_1 (by adjusting $T(z_1)$ and $T(z_2)$ correspondingly and in a budget balanced manner) would be desirable. Setting the derivative of $\tilde{g}(z - T(z), T(z))$ with respect to z to zero, yields $\tilde{g}_c \cdot (1 - T'(z)) + \tilde{g}_{z-c} = 0$ and the optimal tax formula (7). $0 \leq T'(z) \leq 1$ since $\tilde{g}_c \leq 0$ and $\tilde{g}_{z-c} \geq 0$. Note that this is a first-order ordinary nonautonomous differential equation of the form

$$T'(z) = f(z, T(z))$$

with initial condition on $T(0)$ given by the government budget constraint. If $\tilde{g}$ is continuous in both its arguments, so is $f(z, T(z))$ for $z \in [0, \infty)$. Then, by the Cauchy-Peano theorem, a solution $T(z)$ exists, with continuous derivative on $[0, \infty)$. If both $f(z, T(z))$ and $\frac{\partial f(z,T(z))}{\partial z}$ are continuous, then, by the uniqueness theorem of the initial value problem, the solution is unique. ■

A.3 Luck and deserved income multiple equilibria

Proof of Proposition 5: Recall that the optimal linear tax is given by formula (5). Note that with a linear tax redistributed lumpsum, we have $c_i = (1-\tau) \cdot z_i + \tau \cdot Ez$ where Ez is the average of z in the population. Hence, $y_i^l - Ey^l \leq z_i - c_i$ is equivalent to $(Ez - z_i)\tau \leq Ey^l - y_i^l$.

[49] With no income effects on labor supply, this rebate has no further impact on earnings.

Therefore, we can rewrite $\bar{g}$ as:

$$\bar{g} = \frac{\int_i 1((Ez - z_i)\tau \leq Ey^l - y_i^l) \cdot z_i}{\int_i 1((Ez - z_i)\tau \leq Ey^l - y_i^l) \cdot \int_i z_i}$$

At $\tau = 0$, $\bar{g} = \frac{\int_i 1(0 \leq Ey^l - y_i^l) \cdot z_i}{\int_i 1(0 \leq Ey^l - y_i^l) \cdot \int_i z_i}$. Since $Cov(1(0 \leq Ey^l - y_i^l), z_i) < 0$ (as a higher luck income is on average correlated with a higher total income), at $\tau = 0$, $\bar{g} < 1$ and hence the right hand side of the optimal tax formula (5) is positive so that society would like a tax rate τ higher than zero. On the other hand, at $\tau = 0.5$, as long as $e > 1$, as $\bar{g} \geq 0$, the right hand side of (5) is below 0.5 so that society would like a tax rate τ below 0.5. Consequently, between 0 and 0.5 there is one equilibrium tax rate, called the "low tax equilibrium."

Similarly, at $\tau = 0.9$, as long as $e < \bar{g}_0/9$ where $\bar{g}_0$ is the average welfare weight from formula (5) evaluated at $\tau = 0$, then we know that at $\tau = 0.9$, the right hand side of the optimal tax formula in (5) is above 0.9. Hence, by continuity, there has been a point in $[0.5, 0.9]$ where the two sides are equated. That point is the "high tax equilibrium." ■

A.4 Horizontal Equity

Proof of Proposition 6: Suppose $1/(1 + e_2) \geq \tau^*$. Start with the tax system $\tau_1 = \tau_2 = \tau^*$ with τ^* below the revenue maximizing rate $1/(1 + e_2)$ for group 2. Hence, any budget neutral reform with $d\tau_2 < 0$ requires $d\tau_1 > 0$. Given the structure of our weights (that load fully on group 1 which becomes discriminated against), this cannot be desirable either. Naturally, as $e_1 < e_2$, τ^* is also below the revenue maximizing rate $1/(1+e_1)$ for group 1 so that symmetrical reforms $d\tau_2 > 0$ and $d\tau_1 < 0$ are not desirable. Hence, $\tau_1 = \tau_2 = \tau^*$ is an equilibrium.

Let us prove that this equilibrium is unique. Suppose (τ_1, τ_2) is another equilibrium. If $\tau_1 = \tau_2$ then $\tau_1 = \tau_2 > \tau^*$ as τ^* is the smallest uniform rate raising E. Then $d\tau_1 = d\tau_2 < 0$ will typically raise revenue and benefit everybody (as the Laffer curve $\tau \to \tau \cdot (Z_1 + Z_2)$ is single peaked in τ). Hence, we can assume without loss of generality that $\tau_2 < \tau^* < \tau_1$:[50] The equilibrium has horizontal inequity and τ_2, τ_1 bracket τ^*. If not and $\tau_2 < \tau_1 < \tau^*$, then τ^* would not be the smallest uniform τ raising E. If $\tau^* < \tau_2 < \tau_1$ then by singlepeakedness of the Laffer curve in τ_2 decreasing τ_2 (which is above its revenue maximizing rate) would raise revenue and improve everybody's welfare. With $\tau_2 < \tau^* < \tau_1$, it must be the case that $d\tau_2 > 0$ does not raise revenue. If it did, that reform with $d\tau_1 < 0$ would benefit group 1 where all the weight is loaded. Hence, τ_2 is above the revenue maximizing rate $1/(1 + e_2)$ but this contradicts $1/(1 + e_2) \geq \tau^*$.

Suppose $1/(1+e_2) < \tau^*$ and consider the tax system $\tau_2 = 1/(1+e_2)$ and $\tau_1 < \tau^*$ the smallest tax rate such that $\tau_1 \cdot \int_{i \in 1} z_i + \tau_2 \cdot \int_{i \in 2} z_i = E$. τ_2 maximizes tax revenue on group 2. So $d\tau_2 > 0$ requires $d\tau_1 > 0$ to balance budget and is not desirable. $d\tau_2 < 0$ requires $d\tau_1 > 0$ to budget balance and is not desirable as all the weight is loaded on group 1. $d\tau_1 < 0$ with $d\tau_2 = 0$ lowers

[50] The proof in the other case $\tau_2 > \tau^* > \tau_1$ proceeds the same way.

revenue (as τ_1 is the *smallest* tax rate raising E). Hence, this is an equilibrium. Note that $\tau_2 = 1/(1+e_2)$ raises more revenue than $\tau_2 = \tau^*$. Hence, τ_1 does not need to be as high at τ^* to raise (combined with $\tau_2 = 1/(1+e_2)$), total revenue E so that $\tau_1 < \tau^*$.

We can prove that it is unique. First, the equitable tax system $\tau_1 = \tau_2 = \tau^*$ is not an equilibrium because $d\tau_2 < 0$ raises revenue and hence allows $d\tau_1 < 0$ which benefits everybody. Suppose $\tau_2 < \tau_1$ is another equilibrium. Then τ_2 must be revenue maximizing (if not moving in that direction while lowering τ_1 is desirable), then τ_1 must be set as in the proposition. ■

A.5 Poverty Alleviation – Poverty Rate Minimization

Suppose the government cares only about the number of people living in poverty, that is the poverty rate. In that case, the government puts more value in lifting people above the poverty line than helping those substantially below the poverty line. Hence, the social marginal welfare weights are concentrated solely at the poverty threshold $\bar{c}$. Hence $g(c, z; \bar{c}) = 0$ for c below $\bar{c}$ and above $\bar{c}$, and $g(c, z; \bar{c}) = \bar{g}$ for $c = \bar{c}$ ($\bar{g}$ is finite if a positive fraction of individuals bunch at the poverty threshold as we shall see, otherwise $g(c, z; \bar{c})$ would be a Dirac distribution concentrated at $c = \bar{c}$). This implies that $\bar{G}(z) = 0$ for $z \geq \bar{z}$ and $\bar{G}(z) = 1/[1 - H(z)]$ for $z < \bar{z}$.

Proposition 1 *The optimal tax schedule that minimizes the poverty rate is:*

$$T'(z) = \frac{1}{1 + \alpha(z) \cdot e(z)} \quad \text{if} \quad z > \bar{z}$$

$$T'(z) = \frac{-H(z)}{-H(z) + \alpha(z)[1 - H(z)] \cdot e(z)} \quad \text{if} \quad z \leq \bar{z}$$

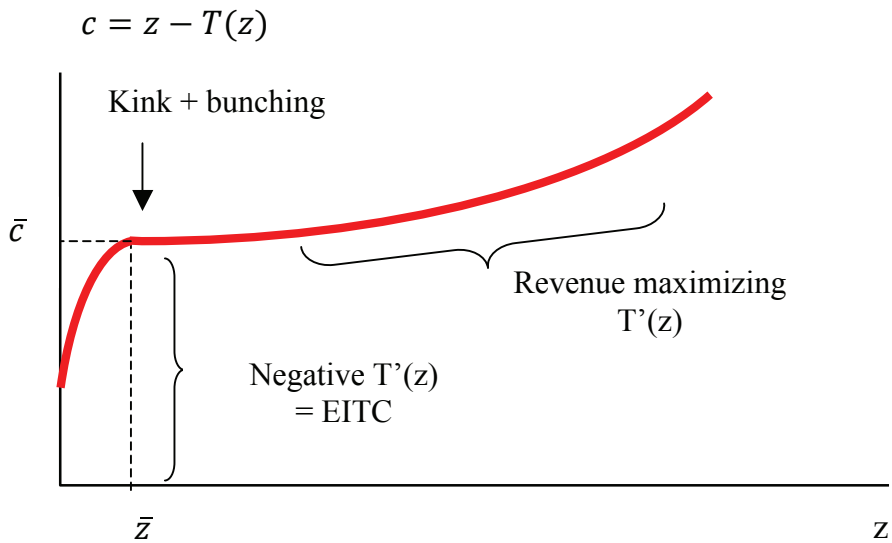
Hence, there is a kink in the optimal tax schedule with bunching at the poverty threshold $\bar{c}$. The marginal tax rate is Rawlsian above the poverty threshold and is negative below the poverty threshold so as to push as many people as possible just above poverty. Hence, the optimum would take the form of an EITC designed so that at the EITC maximum, earnings plus EITC equal the poverty threshold as illustrated in Figure A1.

A.6 Online Survey Description

Our survey was conducted in December 2012 on Amazon's Mechanical Turk service, using a sample of 1100 respondents,[51] all at least 18 years old and US citizens. The full survey is available online at `https://hbs.qualtrics.com/SE/?SID=SV_9mHljmuwqStHDOl`. The first part of the survey asked some background questions, including: gender, age, income, employment status, marital status, children, ethnicity, place of birth, candidate supported in the 2012 election, political views (on a 5-point spectrum ranging from "very conservative" to "very liberal"), and State of residence. The second part of the survey presented people with sliders on which they

[51] A total of 1300 respondents started the survey, out of which 200 dropped out before finishing.

Figure A1: Poverty Rate Minimization



The figure displays the optimal tax schedule in a (pre-tax income z, post-tax income $c = z - T(z)$) plane for poverty rate minimization. The optimal tax schedule resembles an EITC schedule with negative marginal tax rates at the bottom.

could choose the (average) tax rates that they think four different groups should pay (the top 1%, the next 9%, the next 40% and the bottom 50%). The other questions focused on eliciting views on the marginal social welfare weights and are now described in more detail. Parts in italic are verbatim from the survey, as seen by respondents.

Utilitarianism vs. Libertarianism. The question stated: "*Suppose that the government is able to provide some families with a $1,000 tax break. We will now ask you to compare two families at a time and to select the family which you think is most deserving of the $1,000 tax break.*" Then, the pair of families were listed (see right below). The answer options given were: "*Family A is most deserving of the tax break*", "*Family B is most deserving of the tax break*" or "*Both families are equally deserving of the tax break*".

The series shown were:

Series I: (tests utilitarianism)

1) *Family A earns $50,000 per year, pays $14,000 in taxes and hence nets out $36,000.*
   Family B earns $40,000 per year, pays $5,000 in taxes and hence nets out $35,000.
2) *Family A earns $50,000 per year, pays $15,000 in taxes and hence nets out $35,000.*
   Family B earns $40,000 per year, pays $5,000 in taxes and hence nets out $35,000.
3) *Family A earns $50,000 per year, pays $16,000 in taxes and hence nets out $34,000.*
   Family B earns $40,000 per year, pays $5,000 in taxes and hence nets out $35,000.

For purely utilitarian preferences, only net income should matter, so that the utilitarian-

oriented answers should be 1) B is most deserving, 2) Both are equally deserving, 3) A is most deserving. Hence utilitarian preferences should produce a large discontinuity in preferences between A and B when we move from scenario 1) to scenario 2) to scenario 3).

Series II: (tests libertarianism)

1) *Family A earns \$50,000 per year, pays \$11,000 in taxes and hence nets out \$39,000.*
   Family B earns \$40,000 per year, pays \$10,000 in taxes and hence nets out \$30,000.
2) *Family A earns \$50,000 per year, pays \$10,000 in taxes and hence nets out \$40,000.*
   Family B earns \$40,000 per year, pays \$10,000 in taxes and hence nets out \$30,000.
3) *Family A earns \$50,000 per year, pays \$9,000 in taxes and hence nets out \$41,000.*
   Family B earns \$40,000 per year, pays \$10,000 in taxes and hence nets out \$30,000.

For purely libertarian preferences, only the net tax burden should matter, so that the libertarian-oriented answers should be 1) A is most deserving, 2) Both are equally deserving 3) B is most deserving. Hence libertarian preferences should produce a large discontinuity in preferences between A and B when we move from scenario 1) to scenario 2) to scenario 3).

To ensure that respondents did not notice a pattern in those questions, as they might if they were put one next to each other or immediately below each other, we scattered these pairwise comparisons at different points in the survey, in between other questions.

Testing for the weight put on net income vs. taxes paid. In this part of the survey, we created fictitious households, by combining different levels of earnings and taxes paid. Each fictitious household is characterized by a pair (y, τ) where y denotes gross annual income, which could take values in $Y = \{\$10,000; \$25,000; \$50,000; \$100,000; \$200,000; \$500,000; \$1,000,000\}$ and where τ is the tax rate, which could take values in $T = \{5\%, 10\%, 20\%, 30\%, 50\%\}$. All possible combinations of (y, τ) were created for a total of 35 fictitious households. Each respondent was then shown 5 consecutive pairs of fictitious households, randomly drawn from the 35 possible ones (uniformly distributed) and ask to pick the household in each pair which was most deserving of a \$1000 tax break. As an example, a possible draw would be:

"*Which of these two families is most deserving of the \$1,000 tax break?*
Family earns \$100,000 per year, pays \$20,000 in taxes, and hence nets out \$80,000
Family earns \$10,000 per year, pays \$1,000 in taxes, and hence nets out \$9,000"

Test of utilitarianism based on consumption preferences. Utilitarian social preferences lead to the stark conclusion that people who enjoy consumption more should also receive more resources. To test this, we asked respondents:

"Which of the following two individuals do you think is most deserving of a \$1,000 tax break?

- Individual A earns \$50,000 per year, pays \$10,000 in taxes and hence nets out \$40,000. She greatly enjoys spending money, going out to expensive restaurants, or traveling to fancy destinations. She always feels that she has too little money to spend.

- Individual B earns the same amount, $50,000 per year, also pays $10,000 in taxes and hence also nets out $40,000. However, she is a very frugal person who feels that her current income is sufficient to satisfy her needs."

The answer options were again that A is most deserving, B is most deserving, or that both A and B are equally deserving.

Test of Fleurbaey and Maniquet social preferences. To test whether social preferences deem hard-working people more deserving, all else equal, we asked respondents:

"Which of the following two individuals is most deserving of a $1,000 tax break?
- Individual A earns $30,000 per year, by working in two different jobs, 60 hours per week at $10/hour. She pays $6,000 in taxes and nets out $24,000. She is very hard-working but she does not have high-paying jobs so that her wage is low.
- Individual B also earns the same amount, $30,000 per year, by working part-time for 20 hours per week at $30/hour. She also pays $6,000 in taxes and hence nets out $24,000. She has a good wage rate per hour, but she prefers working less and earning less to enjoy other, non-work activities."

The answer options were again that A is most deserving, B is most deserving or that both A and B are equally deserving.

Test of the free loaders model. To test whether the concept of free loaders presented in the main text is relevant for social preferences, we created 4 fictitious individuals and asked people to rank them according to who they deem most deserving. Ties were allowed. The exact question was:

"*We assume now that the government can increase benefits by $1,000 for some recipients of government benefits. Which of the following four individuals is most deserving of the $1,000 increase in benefits? (...)*
- Individual A gets $15,000 per year in Disability Benefits because she cannot work due to a disability and has no other resources.
- Individual B gets $15,000 per year in Unemployment Benefits and has no other resources. She lost her job and has not been able to find a new job even though she has been actively looking for one.
- Individual C gets $15,000 pear year in Unemployment Benefits and has no other resources. She lost her job but has not been looking actively for a new job, because she prefers getting less but not having to work.
- Individual D gets $15,000 per year in Welfare Benefits and Food Stamps and has no other resources. She is not looking for a job actively because she can get by living off those government provided benefits."