<div align="center">

Online Appendix for
"Generalized Social Marginal Welfare Weights
for Optimal Tax Theory"

By Emmanuel Saez and Stefanie Stantcheva

</div>

A   Proofs of Results in the Text

A.1   Local Optimum Approach

In this section, we provide definitions of local optima and derive first and second order conditions.

We focus on money-metric utilities $u_i = c_i - v(z_i; x_i^u, x_i^b)$ (i.e., removing the common concave transformation $u(.)$) so as to express all utility gains or losses in dollar terms.

We start from an initial budget neutral tax schedule T. As in the main text, we denote by $z_i(T)$ the earnings of agent i and by $U_i(T) = z_i(T) - T(z_i(T)) - v(z_i(T); x_i^u, x_i^b)$ the indirect money-metric utility of agent i under the tax system T. We denote by $g_i(T) = g(z_i(T) - T(z_i(T)), z_i(T), x_i^s, x_i^b)$ the generalized social welfare weight under the tax system T.

For any tax schedule $\tilde{T}$, we define the budget function $B(\tilde{T}) = \int_i \tilde{T}(z_i(\tilde{T})) di$ where $\tilde{T}(z_i(\tilde{T}))$ is tax paid by individual i when the tax schedule is $\tilde{T}$. The fact that the initial tax system T is budget neutral implies that $B(T) = 0$. We define the social welfare function $W(\tilde{T}|T)$ from tax system $\tilde{T}$ using the social welfare weights evaluated at T as follows:

$$W(\tilde{T}|T) = \int_i g_i(T) U_i(\tilde{T}) di. \tag{A1}$$

Since the weights $g_i(T)$ are evaluated at the initial tax system T, they are held fixed in this definition, and hence $\tilde{T} \to W(\tilde{T}|T)$ is a standard social welfare function for evaluating tax systems. Because all individual utilities are money-metric, the social marginal welfare weight on individual i is indeed $g_i(T) \cdot u_{c_i} = g_i(T)$. We can define a local tax optimum as follows:

Definition A1 ***Local Optimum.*** *The initial tax system T is a local optimum if and only if there exists a neighborhood of T such that for any budget neutral tax system $\tilde{T}$ (i.e., a tax system such that $B(\tilde{T}) = 0$) in this neighborhood, we have $W(\tilde{T}|T) \leq W(T|T)$.*

Using this formal definition, we can easily prove Proposition 1 from the main text.

As in Proposition 1, consider a perturbed tax system $T + \varepsilon \Delta T$ in direction ΔT that is budget neutral to a first order (according to Definition 2). This implies that the budget function

$\varepsilon \to R(\varepsilon) = B(T + \varepsilon\Delta T)$ is such that (a) $R(0) = 0$ (as the initial T is budget neutral, $R(0) = B(T) = 0$), (b) $R'(0) = 0$ (as the direction of reform ΔT is first order neutral budget, we have $[R(\varepsilon) - R(0)]/\varepsilon = [B(T + \varepsilon\Delta T) - B(T)]/\varepsilon) \to 0$ when $\varepsilon \to 0$). Hence the tax system $\tilde{T}_\varepsilon(z) = T(z) + \varepsilon\Delta T(z) - R(\varepsilon)$ (where $-R(\varepsilon)$ is a lumpsum adjustment) is exactly budget neutral for all ε. Indeed, as there are no income effects, earnings decisions are the same under tax system $T + \varepsilon\Delta T$ and tax system $\tilde{T}_\varepsilon$ so that $B(\tilde{T}_\varepsilon) = B(T + \varepsilon\Delta T) - R(\varepsilon) \equiv 0$.

For ε small enough, the tax system $\tilde{T}_\varepsilon = T + \varepsilon\Delta T - R(\varepsilon)$ is in the neighborhood of T from Definition A1. Because T is a local optimum, we have $W(\tilde{T}_\varepsilon|T) \leq W(T|T)$ for ε small enough and $W(\tilde{T}_{\varepsilon=0}|T) = W(T|T)$. Hence, the function $\varepsilon \to W(T + \varepsilon\Delta T - R(\varepsilon)|T)$ has a local maximum at $\varepsilon = 0$. This implies that the first and second order conditions hold. The first order condition $\frac{dW(T+\varepsilon\Delta T - R(\varepsilon)|T)}{d\varepsilon}|_{\varepsilon=0} = 0$ is exactly equivalent to the condition $\int_i g_i \Delta T(z_i)di = 0$ in the proposition.

The second order condition is $\frac{d^2 W(T+\varepsilon\Delta T - R(\varepsilon)|T)}{d\varepsilon^2}|_{\varepsilon=0} \leq 0$. This second order condition does not have a simple expression but it can be checked on a case by case basis, like in standard optimal tax theory.

A.2   Derivation of the Optimal Tax Formulas using Weights

We show how to derive the optimal nonlinear tax formula (2) and the optimal linear tax formula (3) using the generalized welfare weights approach. In each case, we consider a small budget neutral tax reform. At the optimum, the net welfare effect has to be zero.

Proof of Proposition 2. Optimal non-linear tax.   Consider a small reform $\delta T(z)$ in which the marginal tax rate is increased by $\delta\tau$ in a small band from z to $z + dz$, but left unchanged anywhere else. The reform mechanically collects extra taxes $dz\delta\tau$ from each taxpayer above z. As there are $1 - H(z)$ individuals above z, $dz\delta\tau[1 - H(z)]$ 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 $\delta z = -ez\delta\tau/(1 - T'(z))$ 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 $-dz\delta\tau \cdot h(z)e(z)zT'(z)/(1 - T'(z))$ with $e(z)$ the average elasticity in the small band. Hence, the net revenue collected by the reform is

$$dR = dz\delta\tau \cdot \left[1 - H(z) - h(z) \cdot e(z) \cdot z \cdot \frac{T'(z)}{1 - T'(z)}\right] \tag{A2}$$

This revenue is rebated lumpsum so that the reform is budget neutral. With no income effects, this lumpsum rebate has no labor supply effect on earnings.

What is the effect of the reform on welfare using the generalized welfare weights g_i? The

welfare effect is $\int_i g_i dR di$ for $z_i \leq z$ and $-\int_i g_i(\delta\tau dz - dR)di$ for $z_i > z$. Hence, the net effect on welfare is $dR \cdot \int_i g_i di - \delta\tau dz \int_{\{i:z_i \geq z\}} g_i di$. At the optimum, the net welfare effect is zero. Using the expression for dR above and the fact that $(1 - H(z))\bar{G}(z) = \int_{\{i:z_i \geq z\}} g_i di / \int_i g_i di$, the net welfare effect can be rewritten as

$$dz\delta\tau \cdot \int_i g_i di \cdot \left[1 - H(z) - h(z) \cdot e(z) \cdot z \cdot \frac{T'(z)}{1 - T'(z)}\right] - dz\delta\tau \cdot \int_i g_i di \cdot (1 - H(z)) \cdot \bar{G}(z) = 0 \quad \text{(A3)}$$

Dividing by $dz\delta\tau \cdot \int_i g_i di$ 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 (2).

Discussion of the optimal non-linear tax formula: If average social marginal welfare weights $\bar{g}(z)$ are decreasing in income, $T'(z)$ is always non-negative. To see this, note that by definition $\bar{G}(z) = 1$ when $z = z_{\text{bottom}}$, the bottom of the earnings distribution. Next, $\bar{g}(z)$ decreasing implies that $\bar{G}(z) < 1$ for $z > z_{\text{bottom}}$. Hence, the optimal tax formula then implies that $T'(z) > 0$. The condition that social welfare weights are decreasing in income is always satisfied with standard utilitarian weights ($g_i = u_{c_i}$), or using a standard concave social welfare function $\int_i G(u_i)di$ (with $g_i = G'(u_i)u_{c_i}$). We do not impose this condition *a priori* for all primitive weights g_i in our generalized framework.

Similarly, the zero top tax rate result continues to hold if the income distribution is bounded and average social marginal welfare weights $\bar{g}(z)$ are decreasing in income. The argument is the same as in the Mirrlees model: The top tax rate cannot be positive because otherwise, we could reduce it, which would induce top tax earners to work more (and would not have an adverse effect on revenues above the top earner, since there is no agent earning more). Next, the top tax rate cannot be negative by the argument in the previous paragraph.

Finally, the zero bottom tax rate result continues to hold under the same conditions as in the Mirrlees model, namely, that the bottom earner has strictly positive earnings, in which case the result directly follows from the formula with $\bar{G}(z) = 1$ when $z = z_{\text{bottom}}$.

Proof of Proposition 3. Optimal linear tax. Consider a small reform $\delta\tau$. This increases mechanically tax revenue by $\delta\tau \cdot \int_i z_i di$. By definition of the aggregate elasticity e of $\int_i z_i di$ with respect to $1 - \tau$, this reduces tax revenue through behavioral responses by $-e \cdot \frac{\tau}{1-\tau} \cdot \delta\tau \cdot \int_i z_i di$. Hence, the net effect on revenue is $dR = [1 - e \cdot \frac{\tau}{1-\tau}] \cdot \delta\tau \cdot \int_i z_i di$. This revenue is rebated lumpsum to individuals so that the reform is budget neutral. With no income effects on labor supply, this rebate has no further impact on earnings.

What is the effect of the reform on welfare using the generalized welfare weights g_i? The welfare effect $-\int_i g_i(-dR + z_i \cdot \delta\tau)di$, or, rearranged, $dR \cdot \int_i g_i di - \delta\tau \int_i (z_i \cdot g_i)di$. At the optimum,

this is zero. Using the expression for dR above, this implies:

$$\left[1 - e \cdot \frac{\tau}{1-\tau}\right] \cdot \delta\tau \cdot \int_i z_i di \cdot \int_i g_i di = \delta\tau \cdot \int_i (z_i \cdot g_i) di$$

or equivalently

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

which can easily be re-expressed as the optimal formula (3).

A.3 Proof of Proposition 4

Suppose that we have a non-negative generalized weights function and a local optimum $\tilde{T}(z)$. Consider maximizing the social welfare function SWF:

$$SWF = \int_i (\omega_i \cdot u_i) di \tag{A4}$$

with the Pareto weights such that $\omega_i = g_i / u_{c_i} \geq 0$ where g_i and u_{c_i} are evaluated at the optimum allocation (held fixed in the maximization). For any tax function, consumption is given by $c_i = z_i - T(z_i)$.

We can solve this maximization problem using again a variation approach as in Section A.2 to obtain the same optimal tax formula as in the case for generalized social welfare weights with the individual weights $\omega_i u_{c_i}$. Hence, it is clear that $\tilde{T}(z)$ satisfies the optimal tax formula (2) coming from the first order condition of the maximization of SWF in (A4). ∎

A.4 Taxation with fixed incomes

Proof of Proposition 5: $\tilde{g}(z - T(z), T(z))$ has to be constant with z. Hence, setting the derivative of $\tilde{g}(z - T(z), T(z))$ with respect to z to zero, yields $\tilde{g}_c \cdot (1 - T'(z)) + \tilde{g}_{z-c} \cdot T'(z) = 0$ and the optimal tax formula (4). $0 \leq T'(z) \leq 1$ 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.5 Horizontal Equity

Derivation of the optimal differentiated tax rates:

Individual i belonging to group $m \in \{0,1\}$ chooses z_i to maximize $u_i = z_i \cdot (1 - \tau_m) - v(z_i; x_i^u, x_i^b)$ so that $1 - \tau_m = v_z(z_i; x_i^u, x_i^b)$ and $du_i/d\tau_m = -z_i$ (using the envelope theorem). The government maximizes $\int_{i \in 1} u_i di + \int_{i \in 2} u_i di$ subject to $\tau_1 \cdot \int_{i \in 1} z_i di + \tau_2 \cdot \int_{i \in 2} z_i di \geq E$. Denoting by $Z_m(1 - \tau_m) = \int_{i \in m} z_i di$, the aggregate income in group m as a function of the net-of-tax rate $1 - \tau_m$, we can form the Lagrangian (p is the multiplier):

$$L = \int_{i \in 1} u_i di + \int_{i \in 2} u_i di + p \cdot [\tau_1 \cdot Z_1(1 - \tau_1) + \tau_2 \cdot Z_2(1 - \tau_2) - E]$$

The first order condition in τ_m is:

$$0 = \frac{dL}{d\tau_m} = -\int_{i \in m} z_i di + p \cdot \left[Z_m - \tau_m \frac{dZ_m}{d(1 - \tau_m)} \right] = -Z_m + p \cdot Z_m \left[1 - \frac{\tau_m}{1 - \tau_m} \frac{1 - \tau_m}{Z_m} \frac{dZ_m}{d(1 - \tau_m)} \right],$$

Hence, introducing the elasticity $e_m = \frac{1 - \tau_m}{Z_m} \frac{dZ_m}{d(1 - \tau_m)}$, we have

$$1 = p \cdot \left[1 - \frac{\tau_m}{1 - \tau_m} e_m \right] \quad \text{i.e.,} \quad \frac{\tau_m}{1 - \tau_m} = \frac{1 - 1/p}{e_m},$$

so that, re-arranging, we obtain $\tau_m = (1 - 1/p)/(1 - 1/p + e_m)$ as in the main text. The multiplier p is set so that the government budget constraint is met. Naturally, this requires E to be below the revenue maximizing level that is obtained with $p = \infty$ and the standard revenue maximizing tax rates $\tau_m = 1/(1 + e_m)$.

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 $\delta\tau_2 < 0$ requires $\delta\tau_1 > 0$. 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 $\delta\tau_2 > 0$ and $\delta\tau_1 < 0$ are not desirable. Hence, $\tau_1 = \tau_2 = \tau^*$ is an optimum.

Let us prove that this optimum is unique. Suppose (τ_1, τ_2) is another optimum. If $\tau_1 = \tau_2$ then $\tau_1 = \tau_2 > \tau^*$ as τ^* is the smallest uniform rate raising E. Then $\delta\tau_1 = \delta\tau_2 < 0$ 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 $\tau_2 < \tau^* < \tau_1$:[1] The optimum 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 $\delta\tau_2 > 0$ does not raise revenue.

---

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

If it did, that reform with $\delta\tau_1 < 0$ 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 $\tau_1 \cdot \int_{i \in 1} z_i di + \tau_2 \cdot \int_{i \in 2} z_i di = E$. τ_2 maximizes tax revenue on group 2. So $\delta\tau_2 > 0$ requires $\delta\tau_1 > 0$ to balance budget and is not desirable. $\delta\tau_2 < 0$ requires $\delta\tau_1 > 0$ to budget balance and is not desirable as all the weight is loaded on group 1. $\delta\tau_1 < 0$ with $\delta\tau_2 = 0$ lowers revenue (as τ_1 is the *smallest* tax rate raising E). Hence, this is an optimum. 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 optimum because $\delta\tau_2 < 0$ raises revenue and hence allows $\delta\tau_1 < 0$ which benefits everybody. Suppose $\tau_2 < \tau_1$ is another optimum. 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. ∎

B   Additional Results

B.1   Mapping Pareto weights to generalized welfare weights

Proposition B1 *For any social welfare function of the form $SWF = \int_i (\omega_i \cdot u_i) di$ with $\omega_i \geq 0$ exogenous Pareto weights, there exist generalized social welfare weights with function $g(c, z_i; x_i^s, x_i^b) = \omega_i \cdot u_{c_i}(c_i - v(z_i; x_i^b))$ with $x_i^s = i$, such that the tax system maximizing SWF is an optimum for generalized social welfare weights given the function g.*

The proof is immediate by comparing the first-order conditions for the social welfare maximization to the condition characterizing the optimum with generalized weights in (A3) in Section A.3. Note here that individual identities directly enter the welfare weights, so that, in the notation from the text, $x_i^s = i$. Alternatively, suppose the Pareto weights ω depended on i only through a set of characteristics x_i^s, so that $\omega_i = \omega(x_i^s)$. Then, again, the corresponding generalized weights are directly functions $g(c_i, z_i; x_i^s, x_i^b) = \omega(x_i^s) \cdot u_{c_i}(c_i - v(z_i; x_i^b))$.

B.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.[2] Our framework can capture in a tractable way such social preferences, which differentiate income streams according to their source.[3] These preferences could, under some conditions on the

---

[2]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.

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

income processes, also provide a micro-foundation for generalized social welfare weights $\tilde{g}(c, z-c)$ increasing in $T = z - c$, as presented in Section II.A in the main text.

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. 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:

$$g_i = 1(c_i \leq z_i - y_i^l + Ey^l) \tag{A5}$$

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 her tax confiscates more than the excess of her luck income relative to average luck income. Otherwise, the person receives a zero weight.[4]

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 I, it is necessary to aggregate the individual g_i weights in (A5) 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 optimum 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 (4), which leads to $T(z) = 0$, as society does not want to redistribute deserved income. A real-world example of luck income are health costs. Health costs are effectively negative luck income and the desire to compensate people for them leads to universal health insurance in all advanced economies.

No behavioral responses and unobservable luck income. 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). With unobservable luck income, we make sufficient assumptions that guarantee that any change in total income is partially driven by luck income and partially by deserved income.

Assumption: $y_i^l = \alpha \bar{y}_i + \varepsilon_i$ and $y_i^d = (1 - \alpha)\bar{y}_i - \varepsilon_i$. $\bar{y}_i$ is distributed iid across agents with a density $f_{\bar{y}}(.)$, ε_i is distributed iid across agents with a density $f_\varepsilon(.)$ on $[\underline{\varepsilon}, \bar{\varepsilon}]$, and $0 < \alpha < 1$.

$\bar{y}_i$ is an individual-specific income effect that affects total income: Individuals with high $\bar{y}_i$ have both higher deserved income and higher luck income. On the other hand, α is an economy-wide share factor that determines how much luck income an individual has relative to deserved income for a given total income (say, as a function of the institutional features of the economy). ε_i is a random shock to the split between luck income and deserved income.

---

[4]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 $\tilde{g}(c, z - c)$.

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 5.

If we aggregate the individual weights at each (c, z), we obtain $\tilde{g}(c, z - c) = Prob(y_i^l - Ey^l \leq z_i - c_i | c_i = c, z_i = z)$. Using the expression for luck income and the fact that $z_i = y_i^l + y_i^d = \bar{y}_i$, this can be rewritten as $\tilde{g}(c, z - c) = Prob(\varepsilon_i \leq Ey^l + (1 - \alpha)z_i - c_i | c_i = c, z_i = z)$. By the independence assumption, the distribution of ε conditional on c and z is equal to the unconditional distribution. Hence,

$$\tilde{g}(c, z - c) = \int_{\underline{\varepsilon}}^{(1-\alpha)z - c + Ey^l} f_\varepsilon(x)dx = \int_{\underline{\varepsilon}}^{(1-\alpha)(z - c) - \alpha \cdot c + Ey^l} f_\varepsilon(x)dx$$

.

From the expression above, we obtain: $\frac{\partial \tilde{g}(c, z - c)}{\partial c}\big|_{(z-c)} = -\alpha f_\varepsilon((1 - \alpha)z - c + Ey^l) < 0$. Hence, $\tilde{g}(c, z - c)$ is decreasing in its first argument c.

Next, $\frac{\partial \tilde{g}(c, z - c)}{\partial (z - c)}\big|_c = (1 - \alpha)f_\varepsilon((1 - \alpha)z - c + Ey^l) > 0$. Hence, despite the absence of behavioral effects here, the social weights depend positively on $z - c$, even controlling for c.

As in Proposition 5, 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 unobservable luck income. 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. The discussion here is heuristic. 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. We consider the linear tax case and the rest of the notation is as in Proposition 3. We also assume that individual know their luck income before they make labor supply decisions so that no individual decisions are taken under uncertainty.

Intuitively, there can be multiple locally optimal tax rates if the elasticity e of deserved income with respect to $(1 - \tau)$ is sufficiently high at low tax rates and sufficiently low at high tax rates. This is expected to happen because luck income is inelastic while deserved income is elastic.

To see this, recall that the optimal linear tax is given by formula (3): $\tau = (1 - \bar{g})/(1 - \bar{g} + e)$.

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 $(Ez - z_i)\tau \leq Ey^l - y_i^l$. 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 $\tau = 0$, $\bar{g} = \frac{\int_i 1(0 \leq Ey^l - y_i^l) \cdot z_i}{\int_i 1(0 \leq Ey^l - y_i^l) \cdot \int_i z_i}$. If higher luck income is on average correlated with a higher total income, $Cov(1(0 \leq Ey^l - y_i^l), z_i) < 0$. Then, at $\tau = 0$, $\bar{g} < 1$ and hence the right-hand-side of the optimal tax formula (3) is positive so that society would like a tax rate τ higher than zero. Suppose that at $\tau = 0.5$, $e > 1$ (i.e, deserved income is very elastic and the fraction of deserved income in total income is large at medium tax rate levels such as $\tau = 0.5$). As $\bar{g} \geq 0$, the right hand side of (3) is below 0.5 so that society would like a tax rate τ below 0.5. Consequently, by continuity, there is a tax rate in the interval $[0, 0.5]$ that satisfies the optimal tax formula (3) and also satisfies the second order condition. We call this equilibrium the "low tax optimum."

Similarly, at $\tau = 0.9$, as long as $e < (1 - \bar{g}_{0.9})/9$ where $\bar{g}_{0.9}$ is the average welfare weight from formula (3) evaluated at $\tau = 0.9$ (i.e., at high tax levels, deserved income is small relative to luck income and hence total income is fairly inelastic), then we know that at $\tau = 0.9$, the right hand side of the optimal tax formula in (3) is above 0.9. Hence, by continuity, there is a point in $[0.5, 0.9]$ where the two sides are equated. Note that this equilibrium does not satisfy the second order condition: Just below this equilibrium, decreasing the tax rate is desirable while just above this equilibrium, increasing the tax rate is desirable. Hence, it is not a local optimum.

Furthermore, by the assumption that luck income is exogenous to taxes, we know that at $\tau = 1$, nobody supplies any deserved income and therefore $e = 0$. Hence, $\tau = 1$ is also an equilibrium. Whether this equilibrium is a local maximum or not depends on whether there is an additional equilibrium in $[0.9, 1)$. For instance, suppose that at $\tau = 0.95$, we have $e > (1 - \bar{g}_{0.95})\frac{5}{95}$, where $\bar{g}_{0.95}$ is the average welfare weight from formula (3) evaluated at $\tau = 0.95$. In this case, there is another equilibrium in $[0.9, 0.95]$ which is stable (i.e., satisfies the second order conditions) and, if there are no more equilibria in $[0.95, 1)$, the equilibrium at $\tau = 1$ is unstable (i.e., does not satisfy the second order condition). On the other hand, if there is no additional equilibrium in $[0.9, 1)$, then the equilibrium at $\tau = 1$ satisfies the second order conditions. These additional equilibria are also "high tax optima."

In either case, this heuristic example illustrates the possibility of having multiple equilibria with generalized social welfare weights.

Economies with social preferences favoring hard-earned income over luck income could hence end up in two possible situations. In the low tax optimum, 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 optimum, high taxes lead people to work less, which

9

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.[5]

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

B.3 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 (2) 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 in consumption or to specify a *convex* transformation of $u(.)$ in the social welfare function which undoes the concavity of $u(.)$.

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.[6] 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 (2) 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

---

[5]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 concern for fairness 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.

[6]Atkinson (1975) derives formally the Ralwsian optimal income tax using the maxi-min approach.

most popular model for policy decisions among economists is the median-voter model. Consider one specialization 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 optimum 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 generalized weights where all the weight is concentrated at the median voter.

B.4   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. Yet, let us assume that, in contrast to the analysis of Kanbur, Keen, and Tuomala (1994), the government also wants to respect the Pareto principle.

We can capture such an objective by considering generalized social marginal welfare weights 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 B2 *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. This schedule is indeed closer to the schedule obtained by Kanbur, Keen, and Tuomala (1994) than the poverty gap minimization we considered in the main text. However, in contrast to Kanbur, Keen, and Tuomala (1994), our schedule remains constrained Pareto efficient.

Figure A1: Poverty Rate Minimization

$$c = z - T(z)$$

Kink + bunching

$\bar{c}$

Negative T'(z)
= EITC

Revenue maximizing
T'(z)

$\bar{z}$

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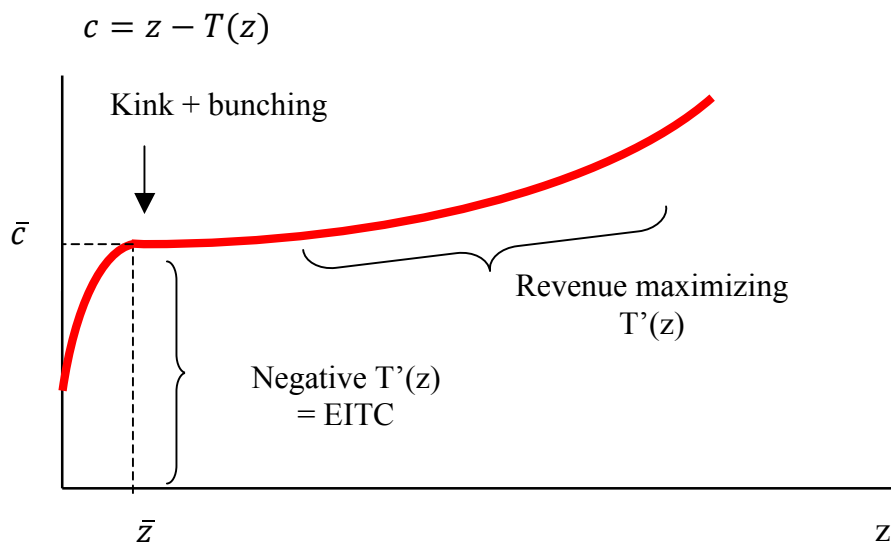
The figure displays the optimal tax schedule in a (pre-tax income z, post-tax income $c = z - T(z)$) plane for poverty rate minimization. The optimal tax schedule resembles an EITC schedule with negative marginal tax rates at the bottom.

B.5 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.[7] 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 the weights using the optimal fair income tax formula. Fleurbaey and Maniquet (2015) provide (independently) a more rigorous and complete connection between the axioms of fair income tax theory and standard optimal income taxation.[8]

---

[7]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, 2015, and the surveys by Kaplow, 2008 and Boadway 2012).

[8]Our approach using formula (2) requires estimating weights by income level. It is of course not always straightforward to derive aggregated weights by income level (see Fleurbaey and Maniquet, 2015 for a discussion

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 (2) 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 (2), this optimal tax system implies that $\bar{G}(z) = 1$ for $0 \le z \le w_{\min}$, i.e., $\int_z^\infty [1 - g(z')]dH(z') = 0$. Differentiating with respect to z, we get $\bar{g}(z) = 1$ for $0 \le z \le w_{\min}$. 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 $w_{\min}$.[9]

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 II.B, 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 $w_i > w_{\min}$, 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 optimum 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.

---

of this important point).

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

C Empirical Testing using Survey Data

The next step in this research agenda is to provide empirical foundations for our theory. There is already 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.[10]

In this section, using a simple online survey with over 1000 participants, we elicit people's preferences for redistribution and their concepts of fairness. Our results confirm that public views on redistribution are inconsistent with standard utilitarianism. We then show how actual elicited social preferences can be mapped into generalized social marginal welfare weights.

The questions of our survey 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 II.A, 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.[11] The complete details of the survey are presented next in Section C.1. 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. The results are presented in Section C.2

C.1 Online Survey Description

Our survey was conducted in December 2012 on Amazon's Mechanical Turk service, using a sample of 1100 respondents,[12] 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

---

[10]See Yaari and Bar-Hillel (1984), Frohlich and Oppenheimer (1992), Cowell and Shokkaert (2001), Fong (2001), Devooght and Schokkaert (2003), Engelmann and Strobel (2004), Ackert, Martinez-Vazquez, and Rider (2007), Gaertner and Schokkaert (2012), Weinzierl (2014), Kuziemko, Norton, Saez, and Stantcheva (2015). 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.

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

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

State of residence. The second part of the survey presented people with sliders on which they 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."

C.2   Results

C.2.1   Qualitative Social Preferences

Table A1 reports preferences for giving a tax break and or a benefit increase across individuals in various scenarios.

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 B.5. 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 II.B 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 II.A with fixed incomes, we analyze whether revealed social marginal welfare weights depend on disposable income and/or taxes paid. Table A2 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 A2 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 II.A.

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 A2 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 A2 provides compelling non-parametric evidence that both taxes and disposable income matter for social marginal welfare weights as we posited in Section II.A.

C.2.2 Quantifying Social Preferences

Table A3 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.[13] 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 A3 shows the implied α and the optimal marginal tax rates in four subsamples.[14] The implied α is between 0.37 and 0.65, so that the implicit optimal marginal tax rates are relatively high, ranging from 61%

---

[13]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%.

[14]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).

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 that of conservatives is much lower at 57%. Liberals put a very small weight on taxes paid relative to disposable income (column 5) while conservative put almost an equal weight on taxes paid relative to disposable income (column 6). Hence, liberals are relatively close to the utilitarian polar case while conservative are about mid-way between the utilitarian and libertarian polar cases.

This simple exercise confirms the results from Table A2 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.

Appendix 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.

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

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.

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

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

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. 2015. "Optimal Taxation Theory and Principles of Fairness," unpublished.

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.

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

Kuziemko, Ilyana, Michael Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015.

"How Elastic are Preferences for Redistribution? Evidence from Randomized Survey Experiments"*American Economic Review* 105(4), 1478–1508.

Lockwood, Benjamin and Matthew Weinzierl. 2015. "De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution." *Journal of Public Economics* 124: 74–80.

Weinzierl, Matthew C. 2014. "The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice." *Journal of Public Economics* 118, 128–142.

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.

Table A1: Revealed Social Preferences

	(1)	(2)	(3)	
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	Harworking = Leisure Lover	Harworking < Leisure Lover	
obs. = 1,121	42.7 %	54.4 %	2.9 %	
	(1.5 %)	(1.5 %)	(0.5 %)	
C. Transfer Recipients and Free Loaders				
	Disabled person unable to work	Unemployed looking for work	Unemployed not looking for work	Welfare recipient not looking for work
obs. = 1,098				
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 to 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 out-of-work 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 person or the welfare recipient not looking for work. For all statistics, standard errors are reported in parentheses below each estimate.

Table A2: 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 % (1.5 %)	54.8 % (1.5 %)	65.2 % (1.4 %)
A=B	38.8 % (1.4 %)	37.3 % (1.4 %)	28.0 % (1.3 %)
A<B	12.4 % (1.0 %)	7.9 % (0.8 %)	6.8 % (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 % (0.8 %)	3.6 % (0.6 %)	3.1 % (0.5 %)
A=B	29.1 % (1.3 %)	40.0 % (1.5 %)	23.7 % (1.3 %)
A<B	63.2 % (1.4 %)	56.4 % (1.5 %)	73.2 % (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 A3: 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 of $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 million} 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 million. 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.