FALL 2019
NEW YORK UNIVERSITY
SCHOOL OF LAW

"Optimal Income Taxation Theory and Principles of Fairness"

Marc Fleurbaey
Woodrow Wilson School of Public and International Affairs

November 5, 2019
Vanderbilt Hall – 202
Time: 4:00 – 5:50 p.m.
Week 10

SCHEDULE FOR FALL 2019 NYU TAX POLICY COLLOQUIUM
(All sessions meet from 4:00-5:50 pm in Vanderbilt 202, NYU Law School)

1. Tuesday, September 3 – Lily Batchelder, NYU Law School.

2. Tuesday, September 10 – Eric Zwick, University of Chicago Booth School of Business.

3. Tuesday, September 17 – Diane Schanzenbach, Northwestern University School of Education and Social Policy.

4. Tuesday, September 24 – Li Liu, International Monetary Fund.

5. Tuesday, October 1 – Daniel Shaviro, NYU Law School.

6. Tuesday, October 8 – Katherine Pratt, Loyola Law School Los Angeles.

7. Tuesday, October 15 – Zachary Liscow, Yale Law School.

8. Tuesday, October 22 – Diane Ring, Boston College Law School.

9. Tuesday, October 29 – John Friedman, Brown University Economics Department.

10. Tuesday, November 5 – Marc Fleurbaey, Princeton University, Woodrow Wilson School.

11. Tuesday, November 12 – Stacie LaPlante, University of Wisconsin School of Business.

12. Tuesday, November 19 – Joseph Bankman, Stanford Law School.

13. Tuesday, November 26 – Deborah Paul, Wachtell, Lipton, Rosen, and Katz.

14. Tuesday, December 3 – Joshua Blank, University of California at Irvine Law School.

Journal of Economic Literature 2018, 56(3), 1029–1079
https://doi.org/10.1257/jel.20171238

Optimal Income Taxation Theory and Principles of Fairness†

Marc Fleurbaey and François Maniquet*

The achievements and limitations of the classical theory of optimal labor-income taxation based on social welfare functions are now well known. Even though utilitarianism still dominates public economics, recent interest has arisen for broadening the normative approach and making room for fairness principles such as desert or responsibility. Fairness principles sometimes provide immediate recommendations about the relative weights to assign to various income ranges, but in general require a careful choice of utility representations embodying the relevant interpersonal comparisons. The main message of this paper is that the traditional tool of welfare economics, the social welfare function framework, is flexible enough to incorporate many approaches, from egalitarianism to libertarianism. (*JEL* D63, H21, H24, J24)

1. *Introduction*

The theory of optimal income taxation has reached maturity and excellent reviews of the field are available (Boadway 2012, Mankiw, Weinzierl, and Yagan 2009, Piketty and Saez 2013b, Salanié 2011). In the classical framework initiated by Mirrlees (1971), the theory studies the maximization of a utilitarian social welfare function by a benevolent planner who only observes the pretax labor income of agents whose wages differ, but whose preferences are identical. The classical framework has been extended in the last two decades in several directions. On the one hand, many contributions have relaxed assumptions in order to take account, for instance, of multidimensional heterogeneity among agents (e.g., Mirrlees 1976, Saez 2001, Choné and Laroque 2010), or to cast the problem in a dynamic framework (e.g., Golosov, Kocherlakota, and Tsyvinski 2003). On the other hand, several authors have recently questioned the ethical foundations of the classical framework and the utilitarian social objective it relies upon. The latter issue is the topic of this paper. We review the different reasons to endorse or to depart from the classical utilitarian objective, and we discuss what policy conclusions the different alternative approaches can deliver.

* Fleurbaey: Princeton University. Maniquet: CORE, Université catholique de Louvain. This paper has benefited from conversations with R. Boadway, S. Coate, E. Saez, and S. Stantcheva, from reactions of participants at the Taxation Theory Conference (Cologne 2014) and the QMUL conference in honor of John Roemer, and comments and suggestions by P. Pestieau, six referees, and the Editor, Steven Durlauf. François Maniquet's work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ ERC n° 269831.

† Go to https://doi.org/10.1257/jel.20171238 to visit the article page and view author disclosure statement(s).

Unease with the classical utilitarian framework is due to internal and external limitations. First, most of the corpus of optimal tax theory assumes that individuals have identical preferences, which is unrealistic. But defining the utilitarian social welfare function for heterogenous utilities requires scaling the utilities and there is no commonly admitted recipe for this. For instance, Boadway (2012) lists the "heterogeneity of individual utility functions" (p. 30) as one of the big challenges for optimal tax theory (along with issues of government commitment, political economy, and behavioral phenomena). "The assumption of identical utility functions is made more for analytical simplicity than for realism. It also finesses one of the key issues in applied normative analysis … which is how to make interpersonal comparisons of welfare" (pp. 30–31).

The problem of interpersonal comparisons of welfare cannot indeed be assumed away for the sake of analytical simplicity. When individuals have the same utility function, the main ethical question that has to be settled is the degree of inequality aversion, over which it is not too difficult to perform a sensitivity analysis spanning the various possible value judgments (from utilitarianism to maximin). This is what optimal tax theory has done very well. In contrast, when individual preferences differ, interpersonal comparisons involve much more difficult questions, which, in philosophy (Rawls 1982) as well as in folk justice (Gaertner and Schokkaert 2012), are generally addressed in terms of fair allocation of resources or opportunities—considerations that move the analysis out of the utilitarian frame.

These considerations connect to the second source of dissatisfaction with the classical framework. It concerns the gap between the normative underpinnings of the theory and the relevant fairness values that seem important in income redistribution. For instance, Weinzierl writes that "conventional theory neglects the diverse normative criteria with which, as extensive evidence has shown, most people evaluate policy" (2012, p. 1). Similarly, Piketty and Saez (2013b) emphasize "the limitations of the standard utilitarian approach" and argue: "While many recent contributions use general Pareto weights[1] to avoid the strong assumptions of the standard utilitarian approach, we feel that the Pareto weight approach is too general to deliver practical policy prescriptions in most cases. Hence, we think that it is important to make progress both on normative theories of justice stating how social welfare weights should be set and on positive analysis of how individual views and beliefs about redistribution are formed" (p. 393). Sheffrin (2013) also argues that folk notions of fairness are ignored in the economic theory of taxation. Among the considerations that are missed by the classical approach, according to these authors, one finds: the libertarian view that the distribution of earnings may deserve some respect; the principle that income inequalities due to differences in preferences or effort are not as problematic as inequalities due to differences in qualifications or social background; and the idea that tagging[2] on the basis of statistical discrimination violates horizontal equity.

The nonutilitarian principles recently revived in the tax literature appear to find support in the empirical literature studying people's values and opinions about redistribution. For instance, Konow (2001) finds support for an "accountability principle" that recommends apportioning rewards to chosen

[1] A (constrained or unconstrained) Pareto-efficient allocation is an extremum on the Pareto frontier for a weighted sum of utilities (and a maximum when the feasible utility set is convex). These weights are called Pareto weights or, in some specific contexts, Negishi weights (Negishi 1972).

[2] Tagging (Akerlof 1978) makes the tax paid by an agent depend on a characteristic that is ethically irrelevant but statistically correlated to some ethically relevant variable, such as the agent's skill.

efforts and to neutralize the impact of external circumstances. Gaertner and Schokkaert (2012) review the empirical social choice literature in which various notions of desert appear relevant in opinion surveys, and, following an experimental approach, Cappelen et al. (2007) and Cappelin, Sørensen, and Tungodden (2010) also find a similar diversity of egalitarian, libertarian, and meritocratic attitudes.[3]

The attempts to make optimal taxation theory sensitive to nonutilitarian ethical values, such as fairness values, have sometimes led scholars to abandon not just the utilitarian criterion but even the classical social welfare function framework. For instance, a certain libertarian tradition has inspired the equal sacrifice approach (Young 1987), which involves looking at departures from market incomes rather than the final distribution of well-being. The idea of abandoning the social welfare function has been recently systematized by Saez and Stantcheva (2016). They propose to apply marginal social welfare weights not to unobservable utility indices but directly to observed earning levels. The weight at each earning level depends on the characteristics of the subpopulation of agents earning that level, and can be inspired by fairness principles, including libertarian notions of desert. In this fashion, Saez and Stantcheva are able to retrieve some of the tax results inspired by fairness principles and to find new results for different principles.

It should be recalled here that when multidimensional heterogeneity is introduced in the taxation problem, there are technical difficulties in identifying the efficient incentive-compatible allocations, due to the bunching of agents of different types at the same income levels. Saez (2001, 2002) proposed to deal with this difficulty by relying on weights applied to earning levels directly, and this technical solution to the multidimensional screening problem is made more attractive if, as Saez and Stantcheva argue, the weights on earnings can be directly derived from ethical principles.

One of our main points, in this paper, is that the classical social welfare function framework is more flexible than commonly thought, and can accommodate a very large set of nonutilitarian values. More specifically, fairness concepts can help solve the interpersonal comparison difficulties that the utilitarian approach faces when agents have different preferences by providing useful selections of suitable individual utility indexes. We devote a significant part of the paper to illustrating this message in the case of the libertarian approach and the resource-egalitarian approach, both of which can be formalized with money-metric utilities that incorporate desert and fairness principles into the objective of the planner. Not all approaches retaining the social welfare function framework are equally successful, however. As argued by Piketty and Saez, the Pareto-weight approach, for instance, is less able than other approaches to deliver recommendations consistent with the fairness values one may wish to capture.

Note that our defense of the social welfare function could even be understood as a defense of the utilitarian approach, for an ecumenical notion of utilitarianism that is flexible about the degree of inequality aversion and the definition of individual utility. Generally, however, utilitarianism is understood as a narrower class of social objectives, with a zero or low aversion to inequality and with a definition of utility that relies on subjective satisfaction or happiness.

In this paper, we mostly confine our attention to the taxation of earnings and the economic literature. There is, of course, an important literature on the taxation of

[3] If one assumes that the prevailing taxes are optimal with respect to social preferences, it may even appear that they reject the Pareto principle (see Bourguignon and Spadaro 2012 for an application to France).

commodities, capital, and inheritance, and we will briefly allude to it at the end. There is also a literature in philosophy (e.g., Murphy and Nagel 2002) and in law (e.g., Zelenak 2006) that discusses the ethical foundations for taxing income and other possible tax bases. It is indeed worth emphasizing that ethical principles may be relevant not only to the design of the income tax, but also to the selection of the tax base. In particular, a controversy (reviewed in Zelenak 2006) rages on the pros and cons of seeking to tax individuals' earning potential rather than actual earnings, and connects to familiar discussions in economics about first-best redistribution. We will briefly refer to these issues when discussing first-best allocations under various approaches.

In the following sections, we begin with a brief description of the main achievements of the classical Mirrlees approach (section 2). We then discuss the various possible interpretations of the concept of utility, viewed as a proxy for well-being (section 3) and examine utilitarianism as an aggregator of well-being levels (section 4). We review various fairness approaches to optimal taxation in section 5: the libertarian approach, Roemer's equality of opportunity, and the resource-egalitarian approach, as well as the luck-desert approach of Saez and Stantcheva (2016). In that section, we also discuss Kaplow and Shavell's (2001) challenge to fairness principles. In the section following that, we examine why it is difficult to incorporate fairness principles in a weighted utilitarian social welfare function (section 6). Next, we analyze the derivation of fair optimal tax and the usefulness of Saez and Stantcheva's approach in terms of marginal social welfare weights (section 7). In section 8, we provide a simple methodology for linking the construction of utility functions with four distinct ethical choices: i) rely on subjective utility or only on fairness principles involving ordinal preferences; ii) insofar as the latter is chosen, reward individual talent or fight inequalities due to unequal skills; iii) insofar as the latter is chosen, prioritize equal tax treatment for identical skills or compensation for unequal skills; and iv) insofar as the latter is chosen, favor individuals with low or high aversion to work. This methodology is meant to be applicable by practitioners who want to be in control of the ethical underpinnings of their choices of utility functions without having to go through arcane axiomatics, and in section 9, we show how the various approaches can be practically used to study tax reforms and to find the optimal tax. We conclude in section 10.

2. *Achievements of the Classical Approach*

Optimal taxation theory studies how to design tax systems that maximize social welfare. Let us begin by defining the main ingredients of the theory formally. As announced, we focus in this paper on the taxation of earnings. There are two goods, labor and consumption, and n agents. A bundle for individual $i \in N = \{1, \ldots, n\}$ is a pair $z_i = (\ell_i, c_i)$, where ℓ_i is labor and c_i consumption. The agents' consumption set X is defined by the conditions $0 \leq \ell_i \leq 1$ and $c_i \geq 0$. The restriction of labor to an interval is not always made in the tax literature, but it will play a role in our analysis of some approaches.

The individuals have two characteristics, their personal utility function over the consumption set and their personal productivity. For every agent $i \in N$, the utility function $U_i : X \to \mathbb{R}$ represents preferences over labor and consumption. We assume that individual utility functions are continuous, quasi-concave, nonincreasing in ℓ, and increasing in c.

The marginal productivity of labor is assumed to be fixed, as with a constant returns to scale technology. Agent i's earning ability is measured by her productivity or hourly wage, denoted w_i, and is measured in consumption units, so that $w_i \geq 0$ is agent i's production when working $\ell_i = 1$ and $y_i = w_i \ell_i$ denotes the agent's earnings (pretax income). Let $V_i(y,c) = U_i(y/w_i, c)$ denote the utility function derived from U_i and representing i's preferences over earnings–consumption bundles.

An allocation is a collection of bundles $z = (z_1, \ldots, z_n)$. A tax function $T : \mathbb{R}_+ \rightarrow \mathbb{R}$ delineates the budget constraint $c = y - T(y)$, which, in terms of labor and consumption, reads $c \leq w_i \ell - T(w_i \ell)$ for all individuals $i \in N$. An allocation is incentive compatible if every agent maximizes his utility in his budget set, or equivalently, if the self-selection constraint is satisfied: for all $i, j \in N$,

$$V_i(y_i, c_i) \geq V_i(y_j, c_j) \quad \text{or} \quad y_j > w_i.$$

An allocation is feasible if $\sum_i T(y_i) \geq G$, where G is an exogenous requirement of government expenditures, or equivalently, $\sum_i c_i \leq \sum_i y_i - G$.

The problem of optimal taxation is to evaluate tax functions and seek the best one under the feasibility and the self-selection constraints. Since Mirrlees (1971), the evaluation of T is derived from a consequentialist evaluation of the allocation(s) z that T generates when every individual makes his choice in his personal budget determined by w_i and T. The evaluation of allocations has to be made with a social-ordering function, which, for every particular population profile $((U_1, \ldots, U_n), (w_1, \ldots, w_n))$, defines a specific ordering (i.e., a complete transitive relation) on the set of allocations X^n.

The classical theory of labor income taxation has been initially developed under two main assumptions. First, the individuals have different productivity levels, but they all have the same preferences over labor–consumption bundles, represented by a single utility function: for all $i \in N$: $U_i = U_0$ or, at least, as shown in Mirrlees (1976, 1986), the heterogeneity of utilities over (y, c) can be described by a single parameter like w_i: $V_i(y,c) = U_0(y/w_i, c)$. Second, the social planner is utilitarian, which means that the social ordering is defined as maximizing the sum of utility levels:

$$\sum_i V_i(y_i, c_i). \tag{1}$$

A more general social welfare function involving transformations of individual utilities has also been considered, but if the social welfare function is additively separable, this is just equivalent to considering various nonlinear rescalings of U_0.

The questions that have been addressed in the optimal tax literature deal with the first-best implications of social welfare maximization, design of second-best[4] tax schemes, and social welfare evaluation of tax reforms. The literature has, in particular, focused on deriving different formulas for the optimal tax rates in the second-best context. These formulas show how marginal tax rates depend on the elasticity of labor supply, the distribution of productivity levels and the shape of the U_0 function (which determines the social marginal value of consumption that the social planner assigns to the different types of agents).

It must be emphasized here that the tax literature has not focused only on the selection of the best tax, but has also explored how to describe the set of efficient incentive-compatible allocations, i.e., the efficient taxes in the second-best context (e.g., Stiglitz 1982, 1987, Guesnerie 1995). This investigation about efficiency has often relied on the weighted variant of the utilitarian social welfare function, which has proved

[4] In the first-best, only the feasibility constraint applies, whereas in the second-best, the self-selection constraint also applies. The first-best is attainable if the government knows individual characteristics, whereas in the second-best, the government only observes earnings (and knows the statistical distribution of characteristics).

to be a versatile tool, and is discussed in more details in section 6.

In the last fifteen years, the theory has been enlarged to consider the more realistic case in which agents differ both in their preferences and their wages. Introducing additional dimensions of heterogeneity makes it considerably more difficult to derive formulas for the optimal tax rates. First, the objective of the planner is much more difficult to define, as it requires one to compare agents with the same productivity but different preferences. Second, the taxation of each income interval influences high-productivity/high-aversion-to-work agents and low-productivity/low-aversion-to-work agents. Determining how much to tax such an interval of incomes is more difficult than when all agents have the same preferences because in the latter case, richer agents also have higher productivity.

Solutions have been found for particular models (see, e.g., Boadway et al. 2002, Jacquet and Van de Gaer 2011, Choné and Laroque 2010). A general solution has also been proposed by Saez (2001, 2002), recently refined and extended in Jacquet and Lehmann (2014). Saez's approach consists of modifying the way the objective of the planner is defined. It is no longer a function of agents' utilities, but a function of agents' incomes. All agents earning the same income, whatever their productivity, receive the same weight, and the objective of the planner is defined in terms of the relative weights that are assigned to sets of people earning different incomes. More details about this approach are provided in section 7. One should also mention that the case of linear taxation is simpler, as shown in Mirrlees (1976) and further studied by Sandmo (1993), showing that the key issue determining whether the optimal tax is progressive or regressive is how preference for leisure affects the marginal utility of consumption.

The weighted-income approach offers a valuable solution to the technical difficulties of optimal tax theory in the presence of heterogenous preferences.[5] Nonetheless, the question of how to make interpersonal utility comparisons and, more specifically, how to compare high-productivity/high-aversion-to-work agents and low-productivity/high-willingness-to-work agents remains complex. This is where fairness considerations can help, as recently advocated by many authors. To prepare the background for such developments, in the next section we go back to the fundamental question of the meaning of utility and its use in optimal tax theory.

3. *What Is Utility?*

The objective of optimal taxation theory is to go beyond the Pareto principle and select among second-best allocations the ones that are better justified from a normative point of view. This requires social evaluation criteria that involve cardinality and/or comparability judgments about individual well-being. Such judgments are embedded in the utilities that enter the computation of social welfare.

There are two main views on utilities. According to the first view, which dominates in the utilitarian tradition, utilities are empirical objects that only need to be measured and can be used as the inputs of a social welfare function, the only ethical issue being the degree of inequality aversion in the function. According to the second view, utilities themselves, not just the social welfare

[5] However, it is primarily a "first-order" approach that does not deal with bunching. Multidimensional heterogeneity is addressed in Saez's (2001) main text, but not in the formal proof of the tax formula. A full formal proof is provided by Jacquet and Lehmann (2014) for separable utility functions and assuming smooth allocations with no bunching (they adopt Wilson's 1993 method of classifying the population into preference types, with single-crossing being satisfied across skills within each type). In its full generality, multidimensional screening remains largely an unsolved problem (Rochet and Stole 2003).

function, are normative indexes that need to be constructed.

The first view has serious weaknesses. One can distinguish two main approaches that adopt this view. In the first approach, utilities refer to subjective self-assessments of well-being. After a long period in which it was believed that only preferences revealed by behavior were a suitable source of data about preferences, the use of surveys in which people express their perceptions has increased. This method has been popularized in the last two decades by the economics of happiness. It builds on answers to survey questions like "Taken all together, how would you say things are these days? Would you say that you are very happy, pretty happy, or not too happy?" There are many versions of this question. A variant relies on answers to questionnaires that request the respondent to decompose their time into a list of activities and, for each of them, to list and evaluate the positive and negative feelings associated with it.[6]

All these approaches are contemporary implementations of ideas that have long been salient in philosophy and economics. Criticisms of these approaches are also well known. The most important, in the context of this paper, comes from political philosophy and was raised by Dworkin (1981a,b), Rawls (1982), and Sen (1985). It says that subjective well-being is not a legitimate argument of a theory of justice. One aspect of the criticism is the "expensive taste" argument. If declaring a lower well-being level only reveals a lower subjective disposition to transform consumption into satisfaction, due to a higher level of aspiration, it does not call for compensation.

Another version of the argument involves adaptation. If declaring a high well-being level only reveals one's ability to adapt to objectively poor conditions, it does not justify a policy failing to address these poor conditions. Decancq, Fleurbaey, and Schokkaert (2015) emphasize that subjective well-being data, by involving heterogenous aspiration levels, produce interpersonal comparisons that may disagree with the comparisons made by the concerned individuals themselves: a highly ambitious high achiever may have a better situation than someone else, as unanimously evaluated by these individuals, and yet have a lower satisfaction level.[7]

Philosophers have suggested the replacement of utilities with other arguments. Dworkin (1981b), in particular, clearly advocates taking the bundles of resources that are assigned to agents as the appropriate argument of a theory of justice. As we will see in section 5, some fairness approaches offer ways to implement an ideal of equality of resources.

This rejection of utilities as an argument of a theory of justice by philosophers seems to echo a similar rejection by people when they are asked to assess allocations. Yaari and Bar-Hillel (1984) have initiated a literature, based on survey questionnaires, dedicated to understanding the ethical principles that guide people's view on just allocations. Summarizing that literature, Gaertner and Schokkaert (2012) write that "the welfarist framework is not sufficient to capture all the intuitions of the respondents. […] Respondents distinguish between needs and tastes and discount subjective beliefs to a large extent. In general, intuitions about distributive justice seem to depend on the context in which the problem is formulated" (pp. 94–5). The same authors also report the

[6] Among many references, see in particular Clark, Frijters, and Shields (2008), Diener, Helliwell, and Kahneman (2010), Di Tella and McCulloch (2006), Dolan and White (2007), Graham (2009), Kahneman, Diener, and Schwarz (1999), Kahneman and Krueger (2006), and Van Praag and Ferrer-i-Carbonell (2008).

[7] A much more extensive discussion of the subjective well-being approach can be found in Fleurbaey and Blanchet (2013, chapter 5).

fact that "issues of responsibility and accountability, of acquired rights and claims, of asymmetry between dividing harms and benefits, are highly relevant to understand real-world opinions" (pp. 137–38). As we will see in section 5, these questions are at the heart of the fairness approaches to optimal taxation.

The second main approach that embraces the idea that utilities are empirical objects that only need to be measured refers to choices under uncertainty. It is well known that rational preferences can be represented by von Neumann–Morgenstern (vNM) utility functions, and such utility functions can be given a cardinal meaning, provided one assumes that risk aversion is a direct translation of preference intensity. This is the assumption that Vickrey (1945) suggested, and Harsanyi (1976) and Mirrlees (1982) endorsed it as well.

There are two main criticisms of this approach. The first criticism opposes the assumption that risk aversion is a measure of intensity of preferences (Roemer 2008). This criticism rejects the view that vNM utility functions can be given an ethically appealing cardinal interpretation.

Even if one accepts the cardinal interpretation, though, vNM utility functions themselves do not provide the comparability that one needs to build a social criterion, and this is especially relevant when one deals with heterogenous preferences. There have been proposals to calibrate the vNM functions, e.g., by letting all individual functions take the same values (0 and 1) at particular points (Hildreth 1953, Dhillon and Mertens 1999, Adler 2012, Sprumont 2013). But it is debatable whether this makes the interpersonal comparisons compelling. Like subjective well-being data, they are vulnerable to the phenomenon that individuals with identical ordinal preferences but different risk attitudes may be ranked by the calibrated vNM functions against their own assessment of their relative situations (even in riskless contexts).

Our conclusion on the literature on empirical measures of well-being is not, however, that they should be rejected. There are authors who are not convinced by the objections against such measures. Some are convinced hedonists and believe that individuals who pursue other goals than happiness are mistaken (Layard 2005, Dolan 2014). Some are more cautiously hoping that these measures are good proxies of well-being and provide meaningful interpersonal comparisons on average (Clark 2016). Our review of the criticisms is meant to substantiate one point: adopting such measures cannot be done without relying on ethical assumptions. They are not neutral and ready-to-use measures.[8]

Once one acknowledges that the choice of a particular utility measure is always strongly value laden, it is a small step to accept the second view and treat utilities as normative constructs. This second view was defended in particular by Atkinson (1995), and it is probably the dominant view among optimal tax theorists. In the context of uniform preferences, Atkinson himself did not advocate relying on subjective well-being measures and instead proposed to choose the least concave utility representation of the preferences of the agents, and then to aggregate them with a more or less inequality-averse aggregator, reflecting the ethical preferences of the social planner. Note that adopting a unique utility function when individuals have identical preferences guarantees that interpersonal comparisons will align with the comparisons made by the individuals themselves—unlike subjective well-being levels based on heterogenous aspiration levels.

Atkinson's calibration is no longer applicable when agents have heterogenous preferences. The least concave utility functions of the individuals are then defined only up to a scaling factor, and are therefore not

[8] The same point was hammered in Robbins's (1935 [1937]) and Bergson's (1938) classical texts.

directly comparable. Additional assumptions are needed to perform adequate interpersonal comparisons.

One of the main ideas that we would like to defend in this paper is that the second view offers valuable ways to accommodate interesting ethical principles about equality and redistribution. It is especially in the case of heterogenous preferences that introducing ethical values may be particularly helpful to guide interpersonal comparisons, and therefore the selection of utility representations. As we will see, it turns out that the traditional concept of money-metric utilities is convenient and surprisingly versatile for the incorporation of libertarian, as well as resource-egalitarian, values. As a matter of fact, even in the case of identical preferences, money-metric utilities, based on libertarian or egalitarian values may help select particular scalings of utilities that are not obviously found when one starts from the least concave representation and wonders about finding a suitable concave transform. Money-metric utilities have not been used much in taxation theory, but have been quite common in other parts of the literature (such as the measurement of living standards across or within different types of households).[9]

4. *Utilitarianism and Just Outcomes*

The social criterion that is classically used in optimal taxation theory is the utilitarian social welfare function that adds up utilities. Several issues have been raised in the literature about this criterion, in relation to certain notions of fairness inspired by egalitarianism, libertarianism, or combinations of both.

The first issue was pointed out by Rawls (1971). Utilitarianism is able to produce the undesirable outcome that a majority imposes an arbitrarily large loss to a minority. To put it differently, utilitarianism is unable to guarantee a safety net to all agents, because an increase in utility of a well-off agent may offset a decrease in utility of a miserable agent, independently of how low the utility of this agent is, and a small utility gain for many well-off individuals may outweigh a large utility loss for the miserable one. This issue, however, is substantially alleviated in inequality-averse variants of utilitarianism.

The second issue was identified by Mirrlees (1974). He proved that in a first-best world, this social welfare function leads to the following surprising result: if all agents have the same preferences, the high-ability agents end up enjoying lower satisfaction levels than the low-ability ones (assuming that leisure is a normal good). This is in sharp contrast with what popular ethical views recommend. One such view is that differences in productive abilities do not justify differences in outcomes. This calls for equalizing satisfaction levels among agents with the same preferences. Another popular view holds that agents own their ability at least partially, so that the high-ability agents should reach a higher satisfaction level.

This issue is related to the common criticism that there is no place in utilitarianism for entitlements and rights. In his critical discussion of the Mirrlees Review, Feldstein (2012) points out the implicit assumption that redistribution is limited only by disincentive effects. How to accommodate libertarian ideas in taxation theory will be discussed in the next section. As we will see in section 9.2, however, certain approaches based on nonutilitarian ethical principles recognizing individual rights can also penalize some or all of the talented individuals in the first-best allocations.

The third issue is emphasized by Piketty and Saez (2013b) and echoes the liberal egalitarian literature initiated by Rawls and Dworkin. Maximizing a sum of utilities, even

[9] See the reviews of this literature in Lewbel (1997) and Chiappori (2016).

weighted,[10] cannot produce the desirable properties that i) utility should be equalized when all agents have the same preferences, an objective which we will refer to later as the *compensation* objective,[11] and ii) laissez-faire should prevail when all agents have the same productivity level, an objective which we will refer to as the *laissez-faire* objective. By "laissez-faire," we mean the imposition of a poll tax $T(y) = G/n$ on all individuals, which is equivalent to the absence of redistribution.

The double goal of equalizing utilities among agents having the same preferences and not redistributing among agents having the same productivity is at the heart of the resource-egalitarian approach, and is also closely connected to the equality of opportunity approach, both of which are reviewed in the next section.

The fourth issue is related to tagging (Mankiw and Weinzierl 2010). It is clear that tagging represents an additional tool in the maximization of a social objective, because it uses relevant correlations between observed and unobserved individual characteristics to better target redistribution. Tagging, though, has been criticized on the ground that it leads to violations of the basic principle of equal treatment of equals. Indeed, if two agents that are identical in all ethically relevant dimensions but differ with respect to the dimension along which people are tagged, they may be treated differently and end up in unequal positions.

If the social criterion is utilitarian, the sum of the utilities will necessarily increase with tagging, but at the cost of these ethically dubious inequalities of treatment between individuals. Moreover, these extra inequalities due to tagging may operate at the expense of the individuals who were already worse off in the absence of tagging.

Let us assume, for instance, that skill is positively correlated to height. Because height is observable, and because tall people should, on average, pay more tax than small agents (as utilitarianism justifies redistributing from richer to poorer agents), the optimal tax scheme would consist first in redistributing a lump-sum amount of money from the tall agents to the small ones and, second, in optimal second-best tax schemes among the small agents and among the tall agents. It is clear that the small unskilled agents will benefit from the tagging, and it is unfortunate that the tall unskilled will suffer as a result.

It is very hard to reject tagging when the objective is maximizing a social welfare function that does not put an absolute priority on respecting horizontal equity. However, the problem is partly alleviated if the criterion is strongly egalitarian, such as a maximin criterion. Whenever such a social criterion increases, by definition this cannot harm the worst off. Concretely, with tagging, the unskilled, whether small or tall, will end up enjoying the same increased satisfaction level. But violations of horizontal equity will still typically happen higher up in the distribution.

Observe how the maximin criterion, applied to any utility indexes, escapes three of the four shortcomings that we just mentioned. It protects the worst off against sacrifices imposed on them to the benefit of well-off individuals. It produces equality in the first best and therefore does not penalize the talented. And it implies respecting horizontal equity among the worst off. However, the maximin criterion has also been severely criticized for granting an absolute priority to the worst off, with devastating consequences for the better off in some circumstances (Arrow 1973, Harsanyi 1976). It is indeed striking when, for instance, the maximin requires choosing a steady state in a growth

10 This holds true unless the weights are endogeneized in a very complex way—see section 6.

11 The word compensation reflects the goal of eliminating the inequalities due to all causes other than preferences.

model in which there exist constant-growth paths in which every generation except the first one is better off than at the maximin allocation. But a main point of this paper is that the choice of the utility index is crucial for this sort of issue. As we will see in the next section, the maximin is actually compatible with a very low degree of redistribution, and even with the most libertarian approach, when rights and entitlements are embedded in the utility index.

Many authors consider that the utilitarian approach is strongly supported by the veil-of-ignorance argument (Harsanyi 1953) and similar arguments such as Harsanyi's (1955) aggregation theorem. In the context of income taxation, one can view the utilitarian criterion as reflecting the ex ante perspective of an individual who could end up at any position in the distribution of wages (Varian 1980 applies this idea to risk bearing on returns to savings). This makes the utilitarian criterion particularly suitable to incorporate the insurance perspective that makes many citizens support redistribution.

Whether the utilitarian criterion is actually appropriate in the context of risk, and whether the veil-of-ignorance idea is a good guide to think about redistribution, are two hotly debated topics in the literature. On the first issue, there are critics who point to the lack of attention of utilitarianism to inequalities, both ex ante (Diamond 1967) and ex post (Broome 1991). On the second issue, Dworkin (2000) has developed a theory of justice around the idea of mimicking what an insurance market would produce if individuals' characteristics were insurable from behind a veil of ignorance. This idea has been criticized for failing to incorporate sufficient attention to inequalities, due to the fact that individuals, ex ante, may want to plan to consume less resources in states with lower marginal utility (e.g., Roemer 1996). Translated into redistribution, this means that individuals born with disabilities that lower their utility should receive less resources if their marginal utility is also lower, and this is the opposite of what typical fairness criteria would recommend.[12]

5. *Fairness Approaches to Optimal Taxation*

In this section, we review the main fairness approaches to optimal taxation. By fairness, we refer to approaches that impose ethical (typically libertarian or egalitarian) requirements on objects other than ordinary utilities. Before we begin this review, though, we need to clarify the relationship between our notion of fairness and the Pareto principle. This can be usefully discussed with Kaplow and Shavell's critique of fairness in the background.

5.1 *Kaplow and Shavell on Fairness*

Kaplow and Shavell (2001, 2002) have argued that any continuous and non-welfarist method of policy assessment violates the Pareto principle, defined as stipulating that an allocation is strictly better than another if everyone is better off in it. In their work, welfarism is defined by the principle of Pareto indifference, according to which society should be indifferent between two allocations if all agents are indifferent between them.[13] Clearly, the Pareto principle, under a suitable continuity condition, implies Pareto indifference. Therefore, under the same continuity condition, a violation of Pareto indifference implies incompatibility with the Pareto principle.

[12] A recent survey on the first issue can be found in Mongin and Pivato (2016). A long discussion of the second issue is in Fleurbaey (2008).

[13] In the theory of social choice (e.g., d'Aspremont and Gevers 2002), welfarism usually combines Pareto indifference with an independence principle, and is defined as the property that not only is the social ordering over allocations derived from a social ordering over utilities, but in addition, the social ordering over utilities remains the same independently of the profile of utilities of the population.

These authors target the normative theories that discriminate among Pareto-indifferent allocations on the basis of fairness principles. A typical example, in the context of taxation, is the principle of horizontal equity stipulating that it is better if people with the same earnings obtain the same final consumption, which we have discussed earlier in relation to tagging. Other examples include the socialist principle that consumption should be proportional to labor, or the libertarian principle according to which the laissez-faire allocation is superior to any other allocation. These authors' critique also aims at the theories of justice that recommend gauging individual advantage not in terms of well-being, but in objective terms such as resources or capabilities.[14]

Kaplow and Shavell argue that fairness principles violating Pareto indifference are therefore harmful and they propose to restrict considerations of fairness to inequality aversion over well-being, which is perfectly compatible with the standard social welfare approach.

As mentioned in the introduction, in this paper, we highlight another way in which at least some fairness principles can remain compatible with the Pareto principle. Fairness principles can indeed guide the selection of the individual utility functions that serve to measure well-being and perform interpersonal comparisons. Insofar as a fairness principle helps picking utility functions among the set of functions that faithfully represent individual preferences, it does not clash with the Pareto principle. Not all fairness principles fall in this category, obviously, and the socialist and libertarian principles mentioned two paragraphs earlier provide examples of non-Paretian approaches.

Surprisingly, however, a good deal of libertarianism can actually be embedded in the choice of the utility functions, as we show in this paper. Many of the fairness approaches that we review in this section can be formulated with a social welfare function and do satisfy the Pareto principle.[15]

5.2 *Libertarianism*

Maybe the most radical departure from the utilitarian approach to taxation is the libertarian view that, absent all kinds of market imperfections, the competitive income distribution is just. The maximization of a social welfare function is replaced by the application of an extended version of the laissez-faire objective: earnings are fair if they reflect the natural differences among people. These differences come from differences in talents, which are rewarded at their marginal productivity, and from differences in preferences, which are rewarded proportionally to labor times. Inequalities due to differences in earning capacities are no longer considered unjust. As Mankiw (2010) puts it, "each person's income reflects the value of what he

[14] See in particular the discussion in Kaplow (2008, chapter 13).

[15] It is easy to be confused about whether such fairness-inspired social orderings should be called welfarist or not, according to the social choice terminology introduced in the previous footnote. Consider a social welfare function $W(u_1^*,\ldots,u_n^*)$ for which u_i^* is a particular representation of individual preferences that is selected out of fairness principles. Obviously, the social ordering over $(u_1^*,\ldots,u_n^*)$ represented by W is welfarist because it does not change when the profile of utilities change. But suppose that one rewrites the same social welfare as a function of empirically given utilities $(u_1,\ldots,u_n)$, e.g., happiness data, so that

$$W(u_1^*,\ldots,u_n^*) = W(\varphi_1^*(u_1),\ldots,\varphi_n^*(u_n)),$$

where the φ_i^* transformations bring the fairness considerations and produce the relevant u_i^*. It is then generally the case that the social ordering over $(u_1,\ldots,u_n)$ is not stable when $(u_1,\ldots,u_n)$ changes, because the transformations φ_i^* have to be changed. For instance, if the selection of u_i^* depends only on individual ordinal preferences, u_i^* does not change when u_i changes without altering the preferences, therefore requiring φ_i^* to change. It is with reference to this observation that Fleurbaey, Tungodden, and Chang (2003) call such fair approaches non-welfarist.

contributed to society's production of goods and services." In a classical paper, Feldstein (1976) had earlier drawn attention to the libertarian viewpoint and its implications for taxation, inspired by Nozick (1974).

The libertarian view is in radical opposition to the pure redistributive objective of labor income taxation, an objective that is shared by utilitarian ethics and the other fairness approaches that we discuss below.

In the presence of market imperfections such as pollution, or public goods such as security, libertarianism is compatible with some taxation of incomes (as explained in Mankiw 2010).[16] The issue of externalities has recently been tackled by Lockwood, Nathanson, and Weyl (2017), who study the Pigouvian taxes that would be needed to curb externalities attached to various professions, in the situation in which professions cannot be taxed directly and one has to estimate the average externalities produced by the various professions present at any given level of earnings. They actually show that being able to directly tax or subsidize occupations would be much more effective than the income tax (even though the policy may still be imperfect because externalities may vary a lot within occupations).

For the case of public goods, there are two main schools of thought. One school advocates taxing according to the benefit received by the tax payers, the other proposes to minimally distort the distribution by imposing equal sacrifice to all. As shown in Weinzierl (2014a), the two approaches may sometimes recommend the same allocation. In general, however, they diverge, with the equal sacrifice approach being more likely to induce progressive taxation, since the same absolute utility loss requires a greater tax for richer taxpayers when marginal utility is decreasing. A classical analysis of equal sacrifice was made by Young (1987).

The benefits approach has been thoroughly studied in the fair allocation literature (classical references are Kaneko 1977, Moulin 1987). In this literature, the distribution of resources is assumed to be fair prior to the production of public goods (which is in line with the libertarian principles of this section) and the question is to compare the gain in well-being obtained by different people with different preferences over the private–public good trade-off. Lindahl equilibria (which treat public goods like private goods but with person-specific prices) have been criticized for being too indeterminate and for failing to satisfy basic fairness conditions (at a Lindahl equilibrium, for instance, identical individuals may end up facing different Lindahl prices). Alternative proposals often rely on the equivalence approach, which consists, in the first-best context, in putting every individual in a situation that is equally good as a benchmark situation (such as enjoying a certain quantity of the public goods for free). While the literature initially focused on the first-best context, the second-best context has been addressed in Maniquet and Sprumont (2004, 2005, 2010) and Fleurbaey and Maniquet (2011a).

But this literature simply ignores earnings, which are left out of the model, and focuses on the contributions to the public good as they relate to individual preferences. Therefore, it is silent about the interaction between income taxation and the funding of public goods. It would be worth exploring how income tax can be used for funding public goods, under such benefit criteria, when there is a known correlation between individual preferences about the private–public good trade-off and individual

[16] The absence of market imperfections is the typical assumption of the Mirrlees optimal tax approach, as the wage rates are assumed to be equal to the productivity of the agents. The Mirrlees framework has of course been extended to take market imperfection into account, such as public goods, or non-competitive labor markets (see reviews in Kaplow 2008 and Boadway 2012).

earnings.[17] In an interesting variant of the model, Weinzierl (2014a) considers the different case in which the public goods affect individuals' private productivity,[18] so that income taxation then appears an even more natural way to apportion contributions to benefits (preferences are assumed to be identical in his model).

Libertarianism, either in the benefit or equal sacrifice variant, does receive some degree of popular support. In a recent survey bearing directly on the ranking of tax schemes, Weinzierl (2014b), for instance, finds that no less than 33 percent of the respondents favor tax schemes in which even the poorest people contribute to the funding of governmental expenditures (20 percent favoring an equal-sacrifice kind of taxation scheme and 13 percent favoring a poll tax). The remaining majority of respondents still favor redistribution toward the poorest, but in a way that is less generous than what the utilitarian or the maximin in utility social welfare functions would recommend.

The main point we want to make here is that the libertarian ethics we just described can still be adapted to the social welfare function framework, at least to some extent. More precisely, we can show that it is possible to construct a social welfare function that achieves the laissez-faire allocation in the absence of market imperfections, and that proposes a benefit-based allocation in the case that a fixed amount of government expenditures has to be collected or public goods have to be funded. The central issue consists in suitably choosing the utility functions representing agents' preferences.

A key concept here is the notion of money-metric utility, due to Samuelson (1974) and which, in this model (after normalizing the price of consumption to 1), can be defined as the value of the expenditure function for a reference wage rate w and the utility level $U_i(z_i)$ (note that in the following formula, as well as later on in the paper, we use t to denote lump-sum transfers):

$$\begin{aligned} m_i(w,z_i) = \ \min\{t \in \mathbb{R} \mid \exists(\ell,c) \in X, \\ c = t + w\ell, \\ U_i(\ell,c) \geq U_i(z_i)\}. \end{aligned}$$

To put it differently, $m_i(w,z_i)$ is the lump-sum transfer (negative if it is a tax) that leaves agent i indifferent between consuming z_i or receiving that lump-sum amount and being free to work at wage w. An important feature of this definition is that wage w plays the role of a parameter of the money-metric index, and need not be the actual wage of agent i. A second feature is that, once w is fixed, the function $m_i(w,\cdot)$ is a numerical representation of i's preferences.[19] For further reference, let the set

$$\{(\ell,c) \in X \mid c = m_i(w,z_i) + w\ell\}$$

be called i's "implicit budget" (at reference wage w). This is the budget with lump-sum transfer and wage rate w that would enable i to obtain utility $U_i(z_i)$. Note that z_i need not belong to that budget.

Consider the following social ordering. It applies an inequality-averse social welfare function W to individual well-being indexes

[17] The classical literature has focused on the validity of the Samuelson rule for public goods in the presence of income taxation (e.g., Christiansen 1981 and Boadway and Keen 1993) and on the marginal cost of public funds (e.g., Gahvari 2006). In a discrete model (finite number of individual types, discrete public good), Bierbrauer (2009) introduces a double heterogeneity in skills and preferences for the public good, but retains a utilitarian objective.

[18] Christiansen (1981) also studies a special case in which the public good is an input.

[19] This is an immediate consequence of the fact that, for any fixed w, the family of sets $\{(\ell,c) \in X \mid c \leq t + w\ell\}$ is nested and continuously increasing in t.

defined as the value of the money-metric utility function at the personal wage rate w_i. This social ordering is then represented by the function

$$W\Big(\big(m_i(w_i, z_i)\big)_{i \in N}\Big).$$

Assuming that $G = 0$, the laissez-faire allocation achieves $m_i(w_i, z_i) = 0$ for all i, and any form of redistribution generates a negative $m_i(w_i, z_i)$ for some i. Moreover, for every feasible allocation the average $m_i(w_i, z_i)$ is non-positive. Therefore, given the inequality aversion of W, the laissez-faire allocation is among the best feasible allocations. Moreover, this ordering is intuitive because it considers that the worst off are those who are in a situation equivalent to suffering the largest lump-sum tax in the population.

Note that the same laissez-faire conclusion is obtained even when W is the maximin criterion

$$W\Big(\big(m_i(w_i, z_i)\big)_{i \in N}\Big) = \min_i m_i(w_i, z_i).$$

The maximin may require that nobody be taxed! This shows that a social ordering based on the maximin aggregator can be compatible with a wide array of redistributive policies. The choice of utility indexes is key.

Let us now look at the optimal allocation according to this maximin ordering if a fixed budget G needs to be collected. Again, the egalitarian objective must yield equality in $m_i(w_i, z_i)$, which is achieved when all agents pay an identical tax G/n and are then free to choose their labor time, being paid at their own wage. The same social ordering that justifies the laissez-faire in the pure redistributive problem therefore recommends the poll tax to finance public expenditures. Note that this is obtained when public expenditures (or their outputs) do not matter directly to individual preferences. A different case is examined below.

Let us come back to the laissez-faire recommendation and study under which conditions it could come out of the maximization of a weighted sum of subjective utilities, for a suitable choice of weights. The laissez-faire allocation being first-best efficient, it is clear (under mild assumptions) that it maximizes a weighted sum of utilities, but the weights depend on the allocation in a complex way. In particular, the weight of an individual depends on characteristics (preferences, productivity) of other individuals. In contrast, the money-metric utility is easy to compute and only depends on the individual's own characteristics. Plugged in to any economy with an arbitrary profile, and any inequality-averse social welfare function, it makes the laissez-faire allocation one of the best. The limitations of weighted utilitarianism will be further discussed in section 6.

To incorporate libertarian values into the social objective, Weinzierl (2014a,b) proposes to take the first-best allocation as a benchmark and maximize a sum of utilities computed in a way that incorporates a cost of deviating from the benchmark. This is tantamount to defining social welfare by the (opposite of the) distance to the desired first-best allocation. This is the most obvious way to extend an allocation rule that selects a particular allocation into a social ordering that ranks all allocations, and can then be maximized in the second-best context. Varian (1976) proposed something similar for the incorporation of the no-envy criterion into a social welfare function. What we suggest here is that the alternative method of incorporating the values that underlie the selection of the first-best goal into the individual utility measures is a less obvious, but quite effective, option.

To see how flexible this alternative methodology is, consider an extension of the model of this paper, in which there is a bundle of public goods g that enters individuals' consumption bundles $z_i = (\ell_i, c_i, g)$. As in Hammond (1994), pick a reference value $\tilde{g}$ and define the money-metric utility

$$m_i(w, z_i) = \min\{t \in \mathbb{R} | \exists (\ell, c) \in X, \quad c = t + w\ell, \quad U_i(\ell, c, \tilde{g}) \geq U_i(z_i)\}.$$

The choice of the reference $\tilde{g}$ is convenient in order to assess the benefit enjoyed by individuals with different preferences. If a low value of $\tilde{g}$ is retained, then the individuals with strong preference for the public goods will be considered pro tanto advantaged by the extra quantity of g that is made available. In contrast, if a value greater than the current g is retained as the reference, the individuals with a strong preference for the public goods will be considered *pro tanto* disadvantaged, i.e., they will have a lower money-metric utility because it subtracts the larger amount of consumption they would be willing to pay for more public goods. This shows that the "benefit" approach can be extended into a "harm" approach if one considers that the current production of public goods is insufficient and harms those who are eager to have more public goods.

If one instead considers Weinzierl's model in which the public goods do not enter utility directly but only affect productivity, one can introduce wage functions $w_i(g)$ and rely on the same notion of money-metric utility, which, in this particular case, reads:

$$m_i(w_i, z_i) = \min\{t \in \mathbb{R} | \exists (\ell, c) \in X, \quad c = t + w_i(\tilde{g})\ell, \quad U_i(\ell, c) \geq U_i(z_i)\}.$$

We will illustrate the implications of this approach for optimal taxation in section 9.

5.3 *Roemer's Theory of Equality of Opportunity*

In contrast to Mankiw, Roemer's (1993, 1998) theory of equality of opportunity is against rewarding individuals for their natural talents, but it retains an idea of desert. This theory is inspired by followers of Rawls and Dworkin in political philosophy, especially Arneson and Cohen (see Arneson 1989, and Cohen 1989). It is also closely connected to the capability approach proposed by Sen (1985). The central idea of the theory is that one should provide individuals with equal or at least equivalent menus of options (called opportunities by some, capabilities by others). This idea naturally implies that the sources of inequalities in individuals' achievements should be divided into two groups. The first group gathers individual characteristics for which agents should not be held responsible. Such characteristics call for compensation, which means that the resulting inequalities in outcomes should be eliminated. They are called the "circumstances" of the individuals, and define the "type" of individuals.

The second group gathers the characteristics for which individuals should be held responsible, typically because individuals control or choose them. They are called "effort variables." Agents should be held responsible for their effort, which means that society should be indifferent to inequalities in agents' outcomes that are caused by such characteristics. This is a key innovation in welfare economics, and it is in sharp contrast with utilitarianism and welfarism as a whole, since the causes of individuals' outcomes play a key role in the evaluation of individual situations.

The social criterion that follows from these principles works as follows. Roemer assumes that individual outcomes are cardinal and comparable. The set of agents is partitioned according to their "genuine" effort—more will be said about this notion in the next paragraph. In each effort group, the worst off are given absolute priority, and social welfare is computed as the average value of outcome for the worst off of all effort subgroups. In other words, social welfare is based on the maximin criterion within effort groups, reflecting the compensation ideal for individuals with identical effort but unequal circumstances; but the utilitarian criterion is applied between effort groups, because there is no concern for inequalities linked to differential effort.

Roemer also advocates a particular way to measure genuine effort. He measures an individual's effort as the percentile of the distribution of outcomes at which this individual stands in the subgroup sharing his circumstances (i.e., the individual's type). The measurement of effort therefore requires partitioning the population by types, and measuring effort within each type by the relative ranks in the distribution of outcome.

Roemer et al. (2003) apply the approach to income taxation. The relevant achievement is assumed to be income. Observe that income is indeed a cardinal and comparable outcome. The set of circumstances is restricted to the level of education of the individuals' parents. The set of efforts is assumed to gather all the characteristics that generate variations in how the influence of parents' education is transformed into income. The tax systems in ten countries are then compared in terms of their ability to equalize the distribution of incomes across types. The optimal linear tax is computed and the tax rate is compared to the average tax rate in the various countries.

One may think of many other applications of Roemer's theory to optimal taxation. In particular, the set of circumstances can be much larger than the parental education level. It would come closer to the classical objective of optimal taxation theory to assume that the circumstances of an agent include her skill. The compensation goal would then be that two agents having different skills but the same effort level should also have the same outcome level. We are then left to define effort and outcome. If income is again retained as the relevant outcome, and individual effort is measured by the agent's percentile in the distribution of his skill group, then the goal becomes the maximization of the average income of the unskilled agents if their distribution of income is first-order stochastically dominated by the income distributions of all other skill groups. This is reminiscent of Besley and Coate's (1995) study of optimal taxation under the goal of minimizing the poverty rate. One worry about such an approach focusing on income is that it is unlikely to satisfy the Pareto principle.[20]

Another possibility, more respectful of individual preferences, would be to take utility as the outcome (assuming there is a comparable measure of utility). The approach would then define effort as the relative rank of an individual in the distribution of utility in his type. If the distribution of utility for unskilled agents is dominated by the distribution of utility for the other types, then the goal is to maximize the average utility of the unskilled agents. This is similar to an approach followed by Boadway et al. (2002) in the special case in which there are two skill levels and two preference types in the economy. In that paper, preferences are assumed to be quasi-linear in leisure, which suggests a natural cardinalization of the preferences.

As Roemer assumes that the relevant outcome is cardinal and comparable, his approach does not solve the difficult question of how to construct utilities. In fact, he explicitly recommends not to apply his approach to utilities and has restricted attention to cases in which the outcomes naturally come in cardinal and comparable

[20] When all individuals enjoy the same circumstances, Roemer's criterion implies maximizing total income.

units, such as incomes. This severely limits the relevance of his approach for optimal taxation conceived as a tool for improving social welfare. But there does not seem to be a fundamental objection against seeking to extend his approach to social welfare by taking some relevant notion of well-being as the outcome.

Responsibility, in Roemer's approach, and desert in the libertarian perspective, seem to follow the same objective, but they are quite different. The difference is best seen if one thinks of an economy in which all agents have the same circumstances, including the same skills. The libertarian approach recommends that agents be rewarded according to their common level of skills, and all income differences then come from different choices, for which no correction is needed. Laissez-faire is then considered fair. In contrast, for this economy, Roemer's equality of opportunity approach recommends indifference about inequalities in outcomes, meaning that the utilitarian criterion should be applied. As a consequence, the optimal policy has no reason to coincide with the laissez-faire (except when utilities are quasi-linear in consumption). The Roemer approach is therefore compatible with income redistribution, even in economies in which all agents have the same earning capacities.

5.4 *The Resource-Egalitarian Approach*

The third fairness approach we survey offers a combination of the previous two approaches. This approach takes inspiration from the resource egalitarian philosophical literature (Rawls, Dworkin) as well as the economic literature on fair allocation theory. It pursues the compensation objective, under the assumption that the characteristics for which individuals are held responsible are their preferences. The compensation principle then requires that agents having the same preferences should also enjoy the same satisfaction level (in the sense of ending up on the same indifference curve).[21] This is clearly a pairwise (hence, stronger) version of the *compensation* objective which, as defined in section 4, dealt only with the case in which all the population has identical preferences.

The approach also retains a responsibility objective, but not Roemer's utilitarian objective. This principle is replaced with a pairwise version of the laissez-faire objective: there should be no redistribution between agents having the same skill level, i.e., they should be submitted to the same degree of redistribution.[22]

Let us note that by combining the pairwise compensation objective with the pairwise laissez-faire objective, this approach offers a solution to Piketty and Saez's criticism of utilitarianism (section 4). Indeed, restricted to economies in which all agents have the same preferences, the pairwise compensation objective boils down to the compensation objective. Restricted to economies in which all agents have the same productive skill, the pairwise laissez-faire objective boils down to recommending the laissez-faire allocation. It is also worth noting that, according to Gaertner and Schokkaert (2012), there is substantial empirical support for this combination of compensation and laissez-faire.[23]

[21] It should be clear that this principle involves only purely ordinal preferences. Enjoying the same satisfaction level, indeed, means that these two agents should consume bundles they deem equivalent. Alternatively, the requirement can be stated by reference to the (ordinal) fairness concept of no-envy, introduced in the formal theory of fair allocation by Kolm (1999 [1972]) and Varian (1974): such agents should not envy each other.

[22] In the literature, the pairwise laissez-faire objective has been variably called the responsibility principle, the natural reward principle, or the liberal reward principle (see Fleurbaey 2008 and Fleurbaey and Maniquet 2011a,b).

[23] However, they also obtain results that are not in accordance with the theoretical literature. For instance, some respondents want to widen the market inequalities, even when they are due to innate talent. A Nietzchean view on redistribution?

In this section, we illustrate the combination of the compensation and the laissez-faire objectives by introducing a new class of social orderings. How to choose a particular ordering among this class, in relation to a variety of ethical principles, is discussed in greater detail in section 8.

The money-metric utility is the key tool, here again. Consider the following social ordering. It applies the *maximin* criterion to individual well-being indexes, which are defined as the value of the money-metric utility function at a common reference wage $\tilde{w}$. This social ordering is then represented by the function[24]

$$\min_{i\in N} m_i(\tilde{w}, z_i).$$

For the moment, let us only assume that the reference wage lies between the lowest and the largest wages observed in the population: $\tilde{w} \in [\min_i w_i, \max_i w_i]$. It, therefore, has to be a function of, at least, the profile of wages in the population. Letting this function remain unspecified, we thus obtain a class of social-ordering functions, rather than a precise one. Let us call this the class of reference-wage egalitarian-equivalent[25] social-ordering functions.

The first point we want to make here is that, in a first-best world, every member of this class of social-ordering functions satisfies the combination of the compensation and laissez-faire objectives discussed in section 4. Consider the case in which all preferences are identical. Pick any common representation of the agents' preferences, U_0. The social ordering that maximizes $\min_{i\in N} U_0(z_i)$ is then exactly the same as every member of the reference-wage egalitarian-equivalent class. Indeed, when preferences are identical, the ranking of individuals in terms of money-metric utilities is then the same as the ranking in terms of utility $U_0(z_i)$, whatever $\tilde{w}$, because $m_i(\tilde{w}, z_i)$ is a numerical representation of the same preferences as U_0, for all $\tilde{w}$. The result that utilities are equalized in a first-best context then follows from the fact that the social ordering is a maximin.

One may worry that when preferences are identical but utility functions differ, picking a common representation or a money-metric utility that only depends on ordinal preferences ignores potentially relevant inequalities in utilities that come from unequal capacities for enjoyment (see, e.g., Boadway 2012, p. 521). In order to examine this issue, two possibilities must be considered.

The first possibility is that the different calibrations of satisfaction simply reflect that some individuals are more difficult to satisfy than others. This directly connects to the discussion of "expensive tastes" and adaptation in section 3. It can be argued that fairness is on the side of well-established approaches that ignore such differences in utilities.

The second possibility is that the capacities for enjoyment reflect internal parameters (metabolism, health, mental health, etc.) that matter to individuals and create real inequalities. This means that the model is incomplete and such internal parameters have to be made explicit, together with individual preferences over them. The bundles can then be denoted (z_i, θ_i), where θ_i denotes the internal parameters. Note that individuals then have three sources of heterogeneity in such an extended model (preferences, wages, and internal parameters).[26] The

[24] The well-being index $m_i(\tilde{w}, z_i)$ is the money-metric utility discussed in Preston and Walker (1999, p. 346) for this same model.

[25] The idea of "egalitarian-equivalence" is due to Pazner and Schmeidler (1978). The expression refers to a social criterion that seeks to achieve an allocation that is Pareto indifferent to an egalitarian allocation. In the case at hand, the egalitarian allocation grants all individuals an equal budget with a lump-sum tax and a wage rate equal to $\tilde{w}$.

[26] Such a triple-heterogeneity model is studied by Valletta (2014). The problem of compensating for unequal internal characteristics is developed at length in a related

extension involves defining money-metric utilities

$$m_i\left(\tilde{w},\tilde{\theta},z_i\right) = \min\Big\{t \in \mathbb{R} | \exists(\ell,c) \in X, \\ c = t + \tilde{w}\ell, \\ U_i\left(\ell,c,\tilde{\theta}\right) \geq U_i(z_i,\theta_i)\Big\}.$$

In conclusion, either way, the worry can be suitably addressed. We continue the discussion, now, under the assumption that agents only differ in productivity and preferences.

Let us come back to the combination of the compensation and laissez-faire objectives and check the laissez-faire side of the picture. When all productivities are equal, $\tilde{w}$ must necessarily equal the common wage, and equality of $m_i\left(\tilde{w},z_i\right)$ is achieved by the laissez-faire allocation (with a poll tax G/n), which is also efficient and incentive compatible, and therefore maximizes the lowest $m_i\left(\tilde{w},z_i\right)$ under the relevant constraints of feasibility and incentive compatibility.

Admittedly, the two cases of identical preferences and equal productivities are very special, and it is important to check if a social-ordering function in the reference-wage egalitarian-equivalent class behaves well in other cases. The main observation is that the compensation property for identical preferences indeed applies to pairs of individuals. Reducing inequalities between two individuals sharing the same preferences is always deemed acceptable for a reference-wage egalitarian-equivalent social ordering, and is even deemed a strict improvement if one considers the leximin variant of such a social ordering.[27] In other words, the pairwise compensation principle (i.e., seek to eliminate all inequalities between pairs of individuals who differ only in their productivities) is fully satisfied.

This compensation effort actually goes beyond the case of individuals with identical preferences. It also applies to cases in which one agent's indifference curve in X lies everywhere above another agent's. In such cases, the money-metric utility of the individual at the lower indifference curve is necessarily lower than the other's. Consequently, the maximin objective recommends to transfer goods from the former to the latter agent. Such a transfer reduces the "inequality in indifference curves" between these two agents.

The same pairwise extension holds for the laissez-faire property, but in a more modest form.[28] It applies to individuals having the same wage, but only when their common wage is equal to $\tilde{w}$. For this special case $w_i = w_j = \tilde{w}$, the counterpart of the laissez-faire ideal is that the equality $m_i\left(\tilde{w},z_i\right) = m_j\left(\tilde{w},z_j\right)$ is achieved in the first best. This means that the two individuals are just as satisfied as they would be by receiving an identical lump-sum transfer or paying an identical lump-sum tax and working at their common wage. For individuals with a common wage $w_i = w_j$ that does not differ too much from $\tilde{w}$, this laissez-faire property will only hold approximately, i.e., the equality $m_i\left(w_i,z_i\right) = m_j\left(w_j,z_j\right)$ will be approximately satisfied.

Therefore, one sees that a reference-wage egalitarian-equivalent social-ordering function consistently seeks to reduce inequalities in indifference curves (whether they belong to the same preferences or not), and in a milder form seeks to avoid allocations that depart too much from the laissez-faire when redistribution is not needed. The

literature (Fleurbaey 2008, Fleurbaey and Maniquet 2011b, 2012).

[27] The leximin extends the maximin lexicographically by considering the very worst-off, then the second worst-off, and so on.

[28] It is impossible to satisfy the two pairwise properties simultaneously (for details, see Fleurbaey and Maniquet 2011a,b).

money-metric utility $m_i(\tilde{w}, z_i)$ makes interpersonal comparisons of resources in an interestingly versatile way to achieve the desired combination of the compensation and laissez-faire objectives: individuals with identical preferences are compared in terms of indifference curves (and this extends to individuals with different preferences but non-crossing indifference curves), and individuals with the reference wage are compared in terms of the lump-sum transfers they receive when tax operates by lump-sum transfers (of which the laissez-faire is a degenerate case).

That these two ways of making interpersonal comparisons can be performed by the same well-being indexes is quite notable. The money-metric utility has generally been considered in the profession as a mere convenience, although it was sometimes presented as more than that. For instance, Deaton and Muellbauer (1980, p. 225) suggest that the money-metric utility reflects "the budget constraints to which the agents are submitted."[29] The money-metric utility respects individual preferences while using an objective measuring rod to compare individual situations. This combination enables it to respect individual preferences not only for intrapersonal comparisons, but also for interpersonal comparisons when individuals unambiguously agree about who has a better situation because indifference curves do not cross. It also enables it to be sensitive to transfers in the case of individuals with the reference wage.

There exist other social-ordering functions that are much stronger with respect to satisfying the pairwise laissez-faire objective (seeking to equalize transfers for all pairs of agents with identical wages) and slightly less strong on compensation (elimination of inequalities in indifference curves for each pair of agents with identical preferences is obtained for a subset of preferences). The basic idea underlying their construction, along with one core example, is developed in section 8. Their implications for taxation are studied in detail in Fleurbaey and Maniquet (2007, 2011a,c).

One might want to question the extreme form of egalitarianism that is adopted through the maximin approach. The literature on fair social orderings (see, in particular, Fleurbaey and Maniquet 2011a, chapter 3), echoing earlier studies of the money-metric utility (Blackorby and Donaldson 1988), shows that mild egalitarian requirements (such as convexity of the social ordering on X^n, or a Pigou–Dalton transfer principle applied to the consumptions of individuals with identical preferences and equal labor) can be satisfied only with an absolute priority for the worst off when the evaluation satisfies informational simplicity requirements. But of course, it is always possible to weaken the egalitarian requirements further in order to obtain a less extreme priority for the worst off.[30]

As we mentioned in the previous subsection, Roemer's theory of equality of opportunity relies on a distinction between circumstances and effort. In the usual optimal income tax model, individuals are characterized by their preferences and their skills. That forces us to put the cut between circumstances and effort between these two characteristics. It would be possible, however, to enrich the model with other elements, such as the influence of family background on preferences, and study

[29] See also Deaton (1980, p. 51): "I believe that practical welfare measurement should be fundamentally based on *opportunities* rather than on their untestable consequences. No government is going to give special treatment to an individual who claims his extra sensibilities require special facilities, at least not without some objective evidence of why money means something different to him than to anyone else." (emphasis added)

[30] See, e.g., Fleurbaey and Maniquet (2005) or Fleurbaey and Tadenuma (2014).

the compensation and laissez-faire objectives with another cut. However, there is an interesting difference between Roemer's approach and this one. In Roemer's perspective, a family background influencing preferences may be a genuine handicap in attaining the relevant outcome (such as income). In the approach described here, there is no comparable outcome and one cannot view an influence on preferences as a handicap in the satisfaction of *these* preferences. Instead, such an influence must be viewed as distorting preferences and implying that the agent's situation should be assessed with "ideal" preferences, i.e., preferences that would be free from the alleged influence. The fairness literature has been reluctant to follow this route because it means dropping the Pareto principle and considering that individual preferences are not fully respectable. This is, however, a route familiar to the literature on behavioral phenomena such as myopia and hyperbolic discounting (e.g., Choi et al. 2003). The relevance of behavioral studies to these issues will be discussed in the conclusion.

5.5 *Luck and Desert*

Saez and Stantcheva (2016, Appendix B2), inspired in particular by Alesina and Angeletos (2005), examine the case in which individual income has two components, the ordinary earnings and a random shock, which for simplicity can be assumed to have a zero mean.[31] Saez and Stantcheva show, in particular, that if only the individuals with net income below their earnings are given an equal positive weight in the social objective, income taxation may have to be greater when the random shock has greater variance relative to net incomes, creating, if earnings are disincentivized by taxation, a reinforcing mechanism that can generate multiple equilibria. Low-tax equilibria incur a lower random shock relative to net incomes, justifiying the low tax, and the converse is true for the high-tax equilibria.

As a matter of fact, the separation between luck and desert is at the core of the literature that has just been reviewed in the previous subsections. The decomposition of gross income into a deserved and an undeserved part can be analyzed easily using Roemer's approach or the resource-egalitarian approach. In both cases, interestingly, the weights on various types of individuals are different from the intuitive ones proposed by Saez and Stantcheva.

In Roemer's approach, the random shock can be added to the circumstances, and if the random shock is uncorrelated with earnings in each type, the partitioning of individuals in terms of "genuine effort" is unchanged. Then, the worst type, when luck is included in the description of type, gathers the individuals with bad luck and otherwise bad circumstances. Roemer's criterion would then advocate maximizing their average outcome.

In the resource-egalitarian approach, one would also treat the random shock as a circumstance to be compensated, and would apply the laissez-faire principle only to the pairs of individuals sharing the same skill and shock. Interestingly, this does not require any change to the index measure. Indeed, the money-metric utility $m_i(\tilde{w}, z_i)$ defined in the previous subsection displays the nice dual property that two individuals with the same preferences, but possibly different skills and/or different luck, should ideally be given final bundles on the same indifference curve, whereas for individuals with wage equal to $\tilde{w}$, the ideal state is to give them lump-sum transfers, canceling their inequalities in luck and

[31] In a classic paper, Varian (1980) studied the case of ex ante identical individuals facing random returns on their savings, and showed how the optimal tax depends on the distribution of shocks. Desert in his model can be viewed as the amount of savings, but all consumers being identical, this dimension of the problem vanishes.

letting them work freely (with no further tax).

The advantage of these two approaches, compared to the intuitive weighting proposed by Saez and Stantcheva, is that it provides a sensible ordering of individuals, which makes it possible to prioritize the very worst off (if the maximin criterion is adopted) or to give a positive weight to all individuals, but with decreasing priority according to their position as measured by $m_i(\tilde{w}, z_i)$.

If one adopts the view that earned income is fully deserved and only the random shock is undeserved, then one can use another utility function, namely $m_i(w_i, z_i)$, which we have shown to be associated with the libertarian approach. As the random shocks are equivalent to redistribution between individuals, any inequality-averse social welfare function applied to the distribution of $m_i(w_i, z_i)$ will seek to undo them via a compensating redistribution. This approach is unlikely to end up giving equal positive weight to those with a net income below their earnings because it will prioritize those with the greater gap.

There is another interesting difference with the weights proposed by Saez and Stantcheva. The weights they propose can generate multiple equilibria because the relative share of the shocks in income is endogenous to the tax. In particular, if the tax is 100 percent, as they note, the posttax earnings are null and all income is undeserved (and has zero elasticity), implying that the optimal tax is indeed 100 percent. In contrast, a social welfare function with $m_i(w_i, z_i)$ is unlikely to generate multiple equilibria and be satisfied with a 100 percent tax. The reason is that it does not treat posttax earnings as deserved when they are strongly distorted by the tax. It treats the tax as being just as bad as a negative shock when it reduces $m_i(w_i, z_i)$. It therefore offers a defense of the principle that individuals deserve to keep their earnings of the laissez-faire allocation, rather than any (distorted) earnings.

6. *Can Weighting Utilities Yield Fair Outcomes?*

If the population comes with a given profile of utility functions $(U_1, \ldots, U_n)$, is it necessary to replace such functions by indexes like $m_i(\tilde{w}, z_i)$ before applying a social welfare function? Couldn't one simply weight these utility functions in a utilitarian sum?

Facing the problem of heterogeneous utilities, the literature[32] has indeed considered weighting them in a utilitarian social welfare function $\sum_{i \in N} \alpha_i U_i(z_i)$. These α_i coefficients are the so-called Pareto weights referred to in the introduction (quoting Piketty and Saez 2013b). It is true that, provided the utility possibility set is convex, every (constrained) efficient allocation can be viewed as optimal for such a weighted utilitarian function. One could therefore imagine seeking the weights α_i that induce the same choice of tax function as, for instance, a reference-wage egalitarian-equivalent social ordering. But this idea does not work well, as we now explain.

Let us consider a second-best world in which only incomes are observable. Assume that the individuals with the lowest wage are sufficiently diverse in preferences so that they span all the labor–consumption preferences of the population. It is then possible to derive the conclusion that, at the socially best allocation, only them should be given a positive weight. This is because, under the incentive constraints, a less productive agent necessarily faces a less favorable budget set than a more productive agent. Therefore, an agent with a high wage will always have a higher well-being index $m_i(\tilde{w}, z_i)$ than an

[32] See, in particular, Boadway et al. (2002), Kaplow (2008), Choné and Laroque (2010).

agent with the same preferences and the lowest wage. The latter agent should then be given full priority over the former, because the objective is a maximin.

The question then becomes: for two agents, say j and k, having different preferences and the lowest wage, how should we determine α_j and α_k? The key point is that their value would have to depend on the whole profile of the population because this profile determines the set of feasible and incentive-compatible allocations, and therefore the exact bundles assigned to these agents in the optimal allocation. That is, once we have identified the second-best optimal allocation for a reference-wage egalitarian-equivalent social ordering, it is possible to compute the corresponding αs. But it is impossible to guess what these weights should be before computing the optimal allocation. Therefore the Pareto-weights methodology cannot help in finding the optimal allocation.

The correct weights, moreover, would be of limited use even if they could be guessed, because the function $\sum_{i\in N}\alpha_i U_i(z_i)$ using these weights is only good at selecting the best allocation. It cannot reliably be used to evaluate suboptimal allocations, for instance in the context of a reform in which both the pre-reform and the post-reform allocations are suboptimal. The evaluation might then go against what the reference-wage egalitarian-equivalent social ordering recommends for such suboptimal allocations, because the individuals' relative marginal utilities may be completely different between the optimal allocation and the suboptimal ones. As a result, new αs 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, again, the values of these αs could only be ascertained after the optimal allocation is identified.

The recent interesting work of Lockwood and Weinzerl (2012) illustrates this difficulty. Following the simplifying method to deal with heterogenous preferences introduced by Mirrlees (1976) and Brett and Weymark (2003), they assume that individual behavioral heterogeneity is one-dimensional. First, preferences are parameterized by a unidimensional number θ_i. Moreover, preferences and skill interact in such a way that all agents with the same $n_i = \theta_i w_i$ are behaviorally indistinguishable and have the same utility $U(y/n, c)$ at the same earning–consumption bundle (y, c).

They rely on a weighted social welfare function

$$\int_0^\infty \alpha_n U(n) f(n)\, dn,$$

where $U(n)$ is a short notation for $U(y(n)/n, c(n))$, the utility of agents i such that $n_i = n$. The marginal social value of consumption $c(n)$ is $\alpha_n g(n)$, where $g(n) = \partial U(n)/\partial c(n)$ is the marginal utility of consumption. They propose to make α_n inversely proportional to

$$E\left[g^{LF}(\theta_i \bar{w})\,|\,n_i = n\right],$$

i.e., the average value, among the agents of (actual) parameter n, of $g(\cdot)$ in the laissez-faire allocation of a hypothetical economy in which all agents have the average wage $\bar{w}$ of the actual economy. When the actual economy already has equal wages for all, the computation of such weights implies that $\alpha_n g^{LF}(n)$ is a constant in n and laissez-faire is an optimal allocation.

An alternative method would start from the individual weights, for each individual i, that would deliver the laissez-faire in the hypothetical economy with equalized wages: $\alpha_i = 1/g^{LF}(\theta_i \bar{w})$. One cannot actually use such weights at the individual level because θ_i is not observed. But, given that the incentive-compatible allocations give the same utility $U(n)$ to agents with the same n, the weighted utilitarian objective can be written:

$$\int_i \alpha_i U_i = \int_0^\infty \alpha_n U(n) f(n)\, dn,$$

for

$$\begin{aligned}\alpha_n &= E[\alpha_i \,|\, n_i = n] \\ &= E\left[1/g^{LF}(\theta_i \overline{w}) \,|\, n_i = n\right].\end{aligned}$$

The evaluation of incentive-compatible allocations according to $\int_0^\infty \alpha_n U(n) f(n)\, dn$ for such weights α_n always coincides with the evaluation that would be made with the correct individual weights α_i.[33]

Whatever the precise formula for the weights, the evaluation of allocations in the *actual* economy is not geared toward the laissez-faire in a systematic way. In particular, the weighted objective may not pursue the pairwise laissez-faire objective in the actual economy, even for agents with average wage. Take two agents, j and k, enjoying the average wage $\overline{w}$ but different θ_j, θ_k. The sum

$$U_j(z_j)/g^{LF}(\theta_j \overline{w}) + U_k(z_j)/g^{LF}(\theta_k \overline{w})$$

does not generally seek equal tax treatment for these two agents in the actual economy when they are far from the laissez-faire bundles they would receive in the hypothetical economy.

Utilitarianism accepts the laissez-faire as optimal (among other allocations, in the first-best context) when individuals have quasi-linear preferences and identical wages. Lockwood and Weinzierl (2015) make use of this insight. The utilities are assumed to be quasi-linear and the weights are set to depend only on w, not on θ.[34] In this case, within any subset of agents with the same wage, the social objective is indifferent to redistribution by lump-sum transfers, and therefore admits equal lump-sum transfers (which represent the "laissez-faire" ideal in this case) as optimal, although it does not strictly prefer such equality to any other distribution of lump-sum transfers with the same sum.

In conclusion, it appears that the replacement of arbitrary utility functions $U_i(z_i)$ by suitable well-being indexes cannot generally be mimicked by a weighting system. Weighted utilitarianism is not an all-purpose tool. Relying on it in order to incorporate the fairness principles that underlie reference-wage egalitarian-equivalent and similar social-ordering functions is possible only if the weights are specific to the allocation that is evaluated and depend on the whole profile of the population. Far from being the simplest amendment to classical welfare economics, weighted utilitarianism seems an arduous detour compared to the direct adoption of well-being indexes such as the no-less-classical money-metric utility.

7. *Weighting Incomes*

An important progress in optimal taxation theory has been recently accomplished by Saez (2001, 2002) and followed up by Saez and Stantcheva (2016). This progress has been made possible by a shift of focus. In Saez' formulation, social preferences are represented by weights that are endogenously assigned to incomes at the contemplated allocation, rather than to utility levels. The underlying rationale is illuminatingly simple. For a social welfare function $\sum_i \alpha_i U_i(z_i)$, a marginal change δT to the

[33] Lockwood and Weinzierl's weights are the harmonic mean of the individual weights α_i, in every n group, instead of the arithmetic mean.

[34] In the quasi-linear case, the method proposed by Lockwood and Weinzierl (2012) produces equal weights for all types.

function T will induce a change in social welfare equal, by the envelope theorem,[35] to

$$(2) \qquad -\sum_i \alpha_i \frac{\partial U_i}{\partial c_i} \delta T(y_i).$$

This expression can be read as a sum over earning levels y of the change in tax $\delta T(y)$ weighted by the total marginal social weight of the subpopulation earning the level y. At a given allocation, one can take the weights $\beta_i = \alpha_i \partial U_i / \partial c_i$ as given and focus on the weighted sum of $\delta T(y)$ over all levels of y. This can be done to evaluate whether a (small) reform is a social improvement, or whether the current allocation is optimal (in which case, no feasible reform yields a positive weighted sum).

The important question raised by Saez and Stantcheva (2016) is whether one can directly guess what the weights should be, or whether one has to deduce them from a social welfare function and a formula like $\beta_i = \alpha_i \partial U_i / \partial c_i$. They provide many examples covering tagging, desert, and fairness, and argue that extending the approach to cases in which the weights are not related to an underlying social welfare function enlarges the set of possible ethical approaches applicable to taxation. Here, we mainly focus on the issue of computing weights that derive from a social welfare function, or at least a social ordering (satisfying transitivity and the Pareto principle). The idea of dropping the social welfare function framework is discussed at the end of the section.

When relying on a social welfare function, a difficulty is that the subpopulation earning y depends on T so that the weights are, in principle, endogenous and depend on the particular allocation under consideration. In this section, we show how the theory of fair social orderings, by relying on specific utility indices, can sometimes help compute these weights.

A simplification comes from the fact that, given that fair social orderings are of the maximin type, a positive weight will only be assigned to the agents exhibiting the lowest value of the well-being index. In a first-best context, of course, these values should be equalized, in which case all agents have a positive weight. In the second-best context of optimal taxation theory, it is quite likely that it will be impossible to equalize well-being levels. In that case, the agents whose well-being is bounded below by incentive compatibility constraints will receive a weight of zero.

In some cases, the objective of maximizing the minimal value of some well-being index enables us to completely determine the weights that should be assigned to incomes. In this section, we present one such case, obtained when the social objective consists of "maximinning" a particular well-being index in the reference-wage egalitarian-equivalent family presented above, $m_i(w_{\min}, z_i)$, where $w_{\min} = \min_{i \in N} w_i$. That is, the well-being of an agent is measured by reference to the lump-sum tax that would leave him indifferent between his actual bundle and freely choosing his labor time, *should he be paid at the minimal wage*.

Let us first examine a reform problem, in which an arbitrary tax scheme prevails but is not optimal. A reform has to be designed, and the tax scheme can only be changed marginally. How should we change it?

The reasoning is illustrated in figure 1. It represents the pretax/after-tax income space. The 45° line represents the relation between pretax and after-tax incomes in the absence of taxation. The tax scheme that we try to evaluate, T, is represented by the

[35] The additional term

$$\sum_i \alpha_i \left(\frac{\partial U_i}{\partial c_i} \big(1 - T'(y_i)\big)\, dy_i + \frac{\partial U_i}{\partial \ell_i} d\ell_i \right)$$

vanishes when either the first-order condition

$$\frac{\partial U_i}{\partial c_i} \big(1 - T'(y_i)\big)\, w_i + \frac{\partial U_i}{\partial \ell_i} = 0$$

or the condition $dy_i = d\ell_i = 0$ (obtained for corner choices at $\ell_i = 0$ or $\ell_i = 1$) holds for all agents.

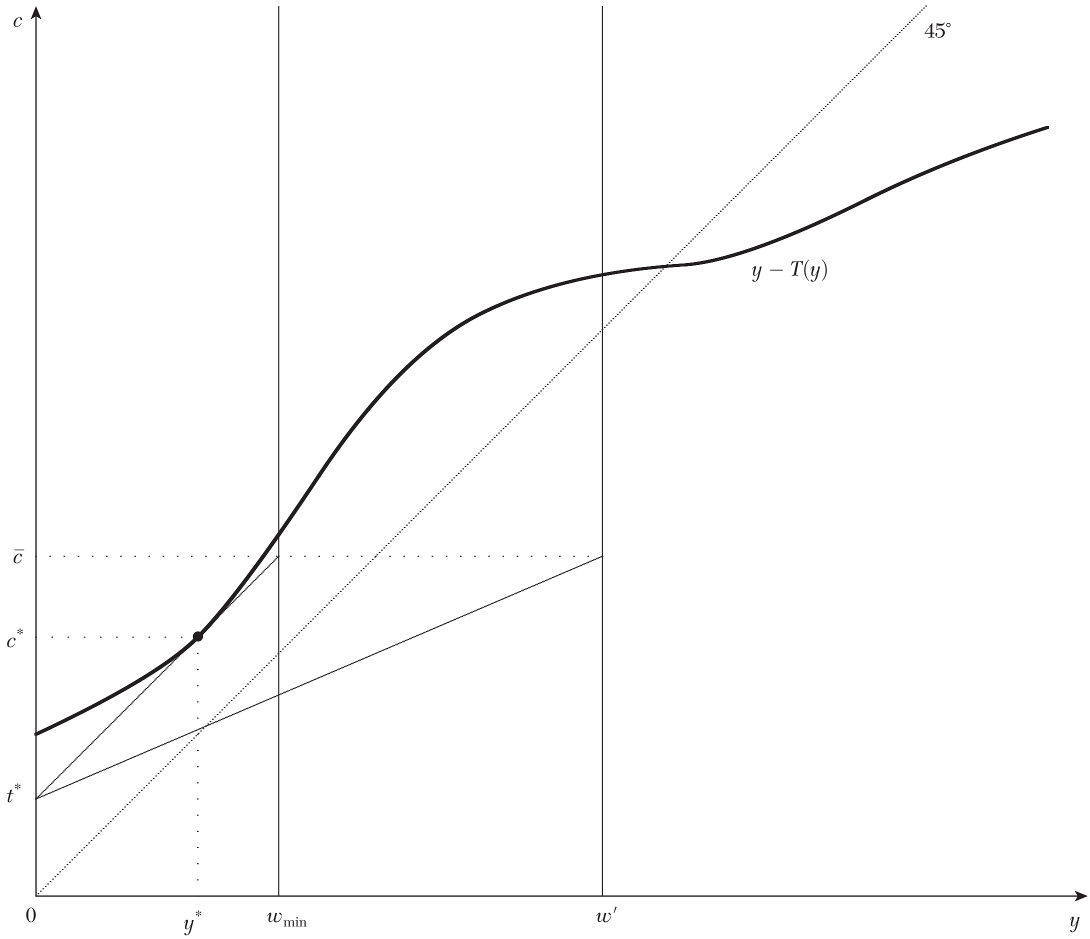

Figure 1. Deriving the Income Weights: The Case of a Reform

corresponding function $c = y - T(y)$ that describes how after-tax incomes c depend on pretax incomes y.[36]

Evaluating a tax scheme requires identifying the agents with the lowest well-being index. It is convenient to do it in two steps. First, the agents with the lowest well-being index need to be identified in each productivity subgroup. Let us begin with agents whose productivity is equal to $w_{\min}$. Given our assumption that labor is bounded ($0 \leq \ell_i \leq 1$), we know that these agents' earnings are in the $[0, w_{\min}]$ interval (remember that w_i also stands for agent i's pretax income, would he work full time).

[36] Note that T exhibits a decreasing marginal tax rate on earnings in the $[0, w_{\min}]$ interval. As a result, the budget set of the low productivity agents is not convex. Nothing in the reasoning here depends on that non-convexity.

Let $y^* \in [0, w_{\min}]$ be the pretax income for which the tax $T(y)$ is maximal over the $[0, w_{\min}]$ interval, or, equivalently, the difference $c - y$ (which is measured in the graph as the vertical distance between the $y - T(y)$ curve and the no-tax line) is minimal. Graphically, y^* is found at the point of tangency between the curve representing $y - T(y)$ and a line segment of slope 1, as drawn on the figure.[37] The intercept associated with that line is denoted t^*.

Let us assume that some agents with minimal productivity, one of them having index i_0, happen to earn y^*.[38] Given the tax scheme, this grants them an after-tax income $c^* = y^* - T(y^*)$.

To formalize the argument, it is convenient to recall the duality relationship between indirect utility and money-metric utility, implying that the indirect utility derived from the implicit budget (defined in section 5.1) is equal to the actual utility level:[39]

$$U_i(z_i) = \max U_i\big(\{(\ell, c) \in X | c \le m_i(\tilde{w}, z_i) + \tilde{w}\ell\}\big)$$
$$= \max V_i\Big(\Big\{(y, c) \in [0, w_i] \times \mathbb{R}_+ | c \le m_i(\tilde{w}, z_i) + \tilde{w}\frac{y}{w_i}\Big\}\Big).$$

Individual i_0's choice reveals that he weakly prefers bundle (y^*, c^*) to all other possible bundles affordable given the tax scheme, but also, by construction of the tangent line segment, to all affordable bundles in the budget $\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le t^* + y\}$. The bundle (y^*, c^*) is the preferred one in both budgets. Therefore, one has

$$V_{i_0}(y^*, c^*)$$
$$= \max V_{i_0}\big(\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le y - T(y)\}\big)$$
$$= \max V_{i_0}\big(\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le t^* + y\}\big),$$

the latter equality implying that $m_{i_0}(w_{\min}, z_{i_0}) = t^*$.

For all other agents with $w_i = w_{\min}$, since

$$\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le t^* + y\}$$
$$\subseteq \{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le y - T(y)\}$$

one always has

$$V_i(y_i, c_i) = \max V_i\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le y - T(y)\}$$
$$\ge \max V_i\{(y, c) \in [0, w_{\min}] \times \mathbb{R}_+ | c \le t^* + y\},$$

implying that $m_i(w_{\min}, z_i) \ge t^*$.

What about individuals for whom $w_i > w_{\min}$? One more line segment is drawn in the figure. It starts from the intercept t^* and ends at consumption level $\bar{c} = t^* + w_{\min}$, for $y = w'$. The slope of this line is $(\bar{c} - t^*)/w'$. Using the equality $\bar{c} - t^* = w_{\min}$, the slope can be expressed as $w_{\min}/w'$.

[37] In this example, y^* is interior to the interval. It need not be the case. If T is convex over $[0, w_{\min}]$, for instance, so that the $y - T(y)$ curve is concave, then either $y^* = 0$, in which case the analysis yields $m_{i_0}(w_{\min}, z_{i_0}) = -T(0)$, or $y^* = w_{\min}$, in which case the analysis yields $m_{i_0}(w_{\min}, z_{i_0}) = -T(w_{\min})$. In both cases, the analysis presented here goes through.

[38] This assumption, actually, can be imposed without loss of generality. Indeed, if no agent earns that income, then the tax scheme is irrelevant at that income level, so that the tax amount can be decreased until it becomes relevant, that is, until one agent is indifferent between her bundle and this new bundle. In that case, we can assume that this agent actually chooses the new bundle. Consequently, the only assumption that is needed is that as soon as a group of agents are willing to earn some income level in the range $[0, w_{\min}]$, there is at least one agent among them who has the lowest productivity $w_{\min}$.

[39] We let $V(B)$ denote the image of the set B by the function V. Therefore $\max V(B)$ is the maximal value obtained by V on set B.

If an agent i of productivity $w_i = w'$ was proposed such a budget, he would reach a utility level

$$\max V_i\Big(\Big\{(y,c) \in [0,w'] \times \mathbb{R}_+ \mid c \leq t^* + \frac{w_{\min}}{w'} y\Big\}\Big),$$

implying that $m_i(w_{\min}, z_i) = t^*$. Now, as is transparent from the figure, any such budget is below the actual budget of this agents, since the function $y - T(y)$ is non-decreasing. As a result, the actual indirect utility cannot be lower, which translates into $m_i(w_{\min}, z_i) \geq t^*$, where z_i denotes the actual bundle of agent i facing T. The inequality is strict if $y - T(y)$ is increasing. Therefore, more generally, all agents having a wage greater than $w_{\min}$ have a strictly larger well-being index than i_0 if $y - T(y)$ is increasing.

In conclusion: assuming that $y - T(y)$ is increasing, one should, at the status quo, assign a positive weight to earning levels like y^*, i.e., corresponding to maximum $T(y)$ for $y \leq w_{\min}$, and a zero weight to all other incomes.

By a simple extension of this reasoning, the result of zero weight for $y > w_{\min}$ also holds under the optimal tax for these social preferences. It is also a small additional step to show that, for the optimal tax, the marginal rate of taxation should be zero on incomes below $w_{\min}$. This gives us a precise formula for the optimal tax scheme. This case is also discussed in Saez and Stantcheva (2016), and here we provide an intuitive explanation for the zero marginal tax result.

The intuition for this result is illustrated in figure 2. Let us consider the tax function T. The marginal rates of taxation differ from zero for incomes below $w_{\min}$. We need to show that this cannot be optimal.

By the same reasoning as above, we can identify the level of the lowest well-being index in the population under T. It is the intercept b of the budget line $c = b + w_{\min}\ell$ that is tangent from below to the budget curve induced by T in the $[0, w_{\min}]$ range of incomes. Formally,

$$b = \min_{y \leq w_{\min}} \big(-T(y)\big).$$

Observe that all agents earning an income of y' or less have a subsidy at least as great as b.

Let us now consider the new tax function T'. It consists in applying a constant amount of subsidy, b, to all income levels lower than y'. Beyond y', T and T' coincide. Compared to T, the new tax scheme T' has two important features.

First, the minimal well-being level remains identical at b. The planner interested in maximinning $m_i(w_{\min}, z_i)$ is, therefore, indifferent between T and T'.

Second, compared to T, the new scheme T' allows the planner to obtain a budget surplus. Indeed, all agents who earn more than y' under T will continue to earn exactly the same income under T'. In particular, the change from T to T' will not induce them to earn less than y', since the lower portion of the budget has become less attractive. Their influence on the budget therefore remains the same.

Agents earning less than y' under T are likely to change their labor time and, therefore, their earning, but the key point is that they will move from an income at which they received a subsidy of at least b to another income at which they receive a subsidy of at most b. That is, no taxpayer will pay less tax under T' than under T, and some of them will pay more. This proves that the planner will now run a budget surplus. By redistributing this surplus to all agents (which can be done by slightly translating the budget curve generated by T' upwards), we can obtain a new allocation that strictly Pareto dominates the previous one, thereby strictly increasing its lowest well-being level above b. This proves that T cannot be optimal. As a

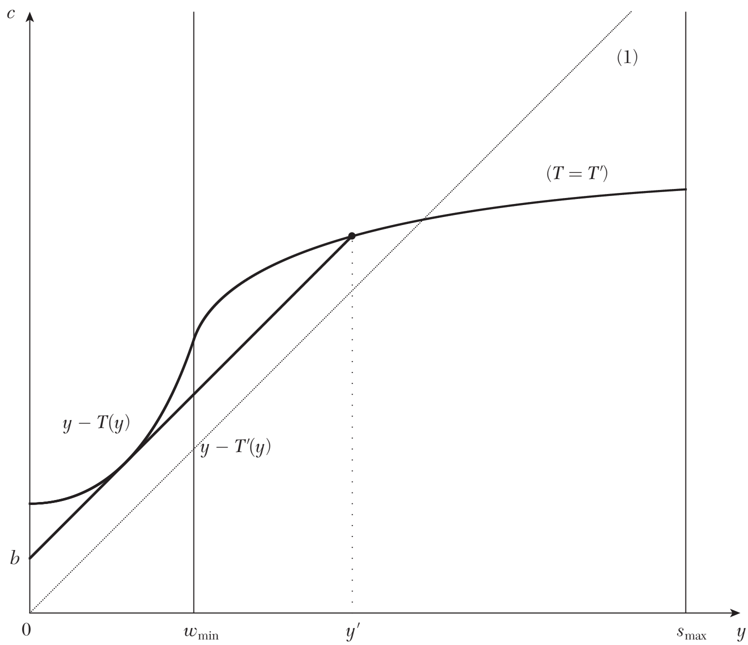

Figure 2. The Optimal Tax Has a Zero Marginal Rate on Low Incomes

result, we need a tax scheme that is flat in the $[0, w_{\min}]$ range of incomes, in which all agents receive the same subsidy: The marginal tax rate is equal to zero in this range.

How should incomes be taxed above $w_{\min}$? As explained earlier, their weight is null and therefore the only objective of taxing those incomes should be to maximize the tax return so as to have as large a subsidy on low incomes as possible (assuming, of course, that after-tax income remains an increasing function of pretax income). This is achieved by applying the Saez (2001) formula with a weight of zero on all incomes above $w_{\min}$, if one assumes that the first-order approach on which the formula relies is valid. In the simple case of no income effect, the marginal taxation rates are a function of the elasticity of the earning supply, the cumulative distribution of the earnings, and its density, which we denote $\epsilon(y), F(y)$, and $f(y)$ respectively. We obtain:[40]

$$\forall y \leq w_{\min}, \qquad T'(y) = 0,$$

$$\forall y \geq w_{\min}, \quad \frac{T'(y)}{1-T'(y)} = \frac{1-F(y)}{\epsilon(y)\, y f(y)}.$$

[40] If the income effect is different from zero, then, the second part of the formula needs to be replaced with

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\epsilon^c(y) y f^*(y)} \int_y^\infty \exp\left[\int_y^{y'} \left(1 - \frac{\epsilon^u(z)}{\epsilon^c(z)}\right) \frac{dz}{z}\right] f(y')\, dy',$$

where ϵ^u and ϵ^c stand for the uncompensated and compensated earning supply elasticity functions, respectively, and f^* is a modified density function. See Saez (2001) for the derivation of this formula, and Jacquet and Lehmann (2014) for a similar formula.

In conclusion, the optimal tax scheme that should be implemented by an egalitarian planner interested in the $m_i(w_{\min}, z_i)$ index of well-being consists of a zero marginal rate on incomes below $w_{\min}$ and a marginal tax rate that follows Saez' formula with zero weights above $w_{\min}$.

There are other cases in which the determination of the weights and optimal tax is less simple. For the different social ordering studied in Fleurbaey and Maniquet (2006), for instance, one can show that the marginal tax rate at a second-best allocation is non-positive on average over low incomes, but the lowest well-being level may be attained by a subset of the low-income agents that is hard to identify, so that the weights at the optimal allocation cannot be easily determined.

In conclusion, this section shows that the weights approach, though useful, does not always provide an easy shortcut, unfortunately, for the determination of the optimal tax. This is because the link between the fairness principles embodied in the social preferences and the weights on earnings is quite complex. Only in some cases involving the maximin criterion can one deduce that some levels of earnings should have zero weight. When the social ordering is not a maximin, the determination of the optimal tax and the associated marginal social welfare weights for incomes is harder.

It may also be worth emphasizing that the social welfare function approach has been introduced by Bergson (1938) and Samuelson (1947) not out of a taste for elegance, but because it is the only way to define social preferences that are both transitive and Paretian.[41] Therefore, a method that directly weights tax changes at the various earning levels is compatible with transitive and Paretian social preferences, and then extendable to the study of nonlocal reforms, only if it relies on the classical framework of the social welfare function. Many reforms to the tax code are not small. Therefore, out of respect for rationality (transitivity) and individual preferences (Pareto), the social welfare function framework seems to provide a safe, even if sometimes complex, toolbox. Fortunately, as we will show in section 9, for an interesting and rather large set of social preferences, there exist extremely simple criteria for the evaluation of reforms—criteria that make it possible not only to easily identify what levels of earnings should be given a positive weight in the evaluation of local reforms, but also make it very easy to assess nonlocal reforms.

8. *Ethical Selection of Utility Indexes*

In the previous sections, we have emphasized the need to carefully select the utility indexes (instead of just weighting them), and have hinted at the wide possibilities offered by the relevant span of indexes. In this section, we analyze this array of possibilities and try to make the underlying ethical choices transparent, so that practitioners can easily connect with the methodology of choosing utility indexes. The theory of fair social orderings has mostly relied on an axiomatic approach. While this is useful to theorists who want to grasp the logical underpinnings of the objects under study, practitioners seek a more direct reading of the meaning of the tools they use. Therefore, here, we ignore the formal aspect of axiomatics and focus on the intuitive meaning of the various features on an index of well-being. Our analysis here echoes the comprehensive review of indexes in Preston and Walker (1999), where the underlying normative principles were not made explicit, and an analysis in Decoster

[41] This claim is formally true only if transitivity is logically strengthened into being representable (by a function). Transitive Paretian social preferences may or may not be representable by a function. The social welfare function approach is here meant to include all such social preferences.

and Haan (2015), where three main examples of indexes are discussed together with their ethical meaning.

8.1 *Four Choices*

We will focus on four ethical choices that guide the selection of a well-being index in the taxation model:

1. Does one trust subjective utility or rely only on ordinal preferences?
2. Does one seek to reduce inequalities due to unequal skills or consider that individuals deserve their own productivity?
3. Does one prioritize the inequalities due to unequal skills or the inequalities of tax treatment between equally skilled individuals?
4. Does one want to pay special attention to the individuals with high or low aversion to work?

While the first two questions are obvious and relate to issues already discussed, the last two are less obviously relevant and their importance will be explained below.

8.1.1 *Does One Trust Subjective Utility or Rely Only on Ordinal Preferences?*

As explained in section 3, indexes that are constructed on the basis of individuals' ordinal preferences are less vulnerable to expensive tastes and the adaptation problem than subjective declarations of utility or satisfaction. Moreover, subjective declarations may contradict the individuals' own interpersonal comparisons, and the latter are especially compelling when indifference curves do not cross. An individual may be on a higher indifference curve but declare a lower satisfaction than another, simply because individual self-assessments rely on heterogeneous standards.

It may be objected that utility takes account not only of standards of satisfaction, but also of other aspects of the individual's situation than the labor–consumption bundle that is the focus of the taxation model. This is an important objection, and it calls for embedding the model into a larger space in which the relevant aspects of life that individuals care about are taken into account. Truly enough, this objection falls short of providing a reason to take subjective utility at face value. Nonetheless, it does raise a serious question when the model is not enlarged, because an approach that only looks at ordinal preference over labor–consumption bundles may miss important dimensions of inequality.

It is commonly believed that no interpersonal comparisons can be made on the basis of individual ordinal noncomparable preferences. In the rest of this section, we focus on indexes that are based on ordinal noncomparable preferences in order to show the wide array of possibilities offered by this informational basis.

8.1.2 *Does One Seek to Reduce Inequalities due to Unequal Skills or Consider That Individuals Deserve the Fruits of Their Talents?*

We have seen in section 5 that the money-metric utilities $m_i(\tilde{w}, z_i)$, fed into an inequality-averse social welfare function, do seek to eliminate inequalities due to skills among individuals having the same preferences, whereas the different money-metric utilities $m_i(w_i, z_i)$ embody the libertarian goal of letting the skilled individuals enjoy their advantage.

Interestingly, one could explore a compromise view in which one would use the indexes $m_i\big(\lambda\tilde{w} + (1-\lambda)\,w_i, z_i\big)$ in order to let individuals enjoy their skills to some extent $(1-\lambda)$ and limit inequalities due to skills to the complementary extent (λ).

More importantly, one can also generalize the indexes and observe that a money-metric utility really defines a budget, rather than just a number. One can then choose what part of the budget to consider for the comparison across money-metric utilities. This, of course, is of no consequence for money-metric utilities of the $m_i(\tilde{w}, z_i)$ sort, because all the implicit budgets are parallel. But for the $m_i(w_i, z_i)$ indexes, it matters a lot, because their slope in the (ℓ, c) space is w_i, so that the budget lines often cross.

A simple generalization of the $m_i(w_i, z_i)$ indexes, therefore, consists in picking a reference value $\tilde{\ell}$ for labor and evaluating how much one would consume with the implicit budget if one worked $\tilde{\ell}$ hours. This is computed as the index $m_i(w_i, z_i) + w_i \tilde{\ell}$. Such an index has been advocated at length in Kolm (2004). It is grounded in the idea that everyone should consider $\tilde{\ell}$ as the fraction of one's time due to the community, and can fully enjoy the benefits of the remainder. When $\tilde{\ell} = 0$, one obtains the standard libertarian view that everyone fully owns one's talents.

When this new index is equal across individuals, their implicit budgets cross at the point where $\ell = \tilde{\ell}$. To sum up this discussion, we obtain the general index

$$m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell},$$

where $\tilde{w}_i = \lambda \tilde{w} + (1 - \lambda) w_i$.

The construction of the value of this index for a general bundle z_i and utility function U_i is illustrated in figure 3 (slopes of budget lines are noted between parentheses below the line). With such an index, there are two ways to seek redistribution across individuals with unequal skills. One way is to let the personalized reference wage $\tilde{w}_i$ be equal across individuals ($\lambda = 1$), yielding the index $m_i(\tilde{w}, z_i) + \tilde{w}\tilde{\ell}$ which is, up to a constant, equivalent to $m_i(\tilde{w}, z_i)$. Another way to adopt a redistributive attitude is to pick a large value for $\tilde{\ell}$, because one then seeks to make the implicit budgets for the skilled agents low compared to those of the less-skilled agents. When $\tilde{\ell} = 1$, one seeks to equalize the full incomes corresponding to these implicit budgets, which is extreme[42] because the implicit budgets for greater skills are then dominated by the implicit budgets for lower skills (except at $\ell = 1$, where they meet).

For ease of reference, we will call "egalitarian-equivalent" the case in which $\lambda = 1$, i.e., $\tilde{w}_i = \tilde{w}$, and "libertarian" the case in which $\lambda = 0$, i.e., $\tilde{w}_i = w_i$, even though, when $\tilde{\ell} > 0$, a substantial amount of redistribution can ensue (only when $\tilde{\ell} = 0$ does one have "pure" libertarianism).

A further generalization (explored in Fleurbaey and Maniquet 1996 and Fleurbaey 2008) would not focus on a particular labor reference $\tilde{\ell}$, and would instead pick a reference preference ordering and apply a corresponding indirect utility function to the various implicit budgets.[43] In other words, budgets would be compared by the indifference curves of the reference preference relation that are tangent to the budgets.

There are, thus, two ways of introducing a redistributive attitude in the indexes $m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell}$, where $\tilde{w}_i = \lambda \tilde{w} + (1 - \lambda) w_i$: let $\lambda \to 1$ or let $\tilde{\ell} \to 1$. A natural question here is whether taking one route or the other makes a difference. It does, and this is where the next question becomes relevant.

8.1.3 *Does One Prioritize the Inequalities due to Unequal Skills or the Inequalities of Tax Treatment between Equally Skilled Individuals?*

Compare the properties of $m_i(\tilde{w}, z_i)$ and $m_i(w_i, z_i) + w_i \tilde{\ell}$. The former always considers that an individual on a dominated indifference curve is worse off, and therefore

[42] See section 9.2 for further discussion.

[43] The reference $\tilde{\ell}$ corresponds to the case in which the reference preferences always choose a labor time equal to $\tilde{\ell}$ whatever the actual productivity w_i. The generalization amounts to assuming that $\tilde{\ell}$ may itself depend on $\tilde{w}_i$, in which case it can only be an increasing function of $\tilde{w}_i$, because the compensated labor supply cannot but increase in the wage.

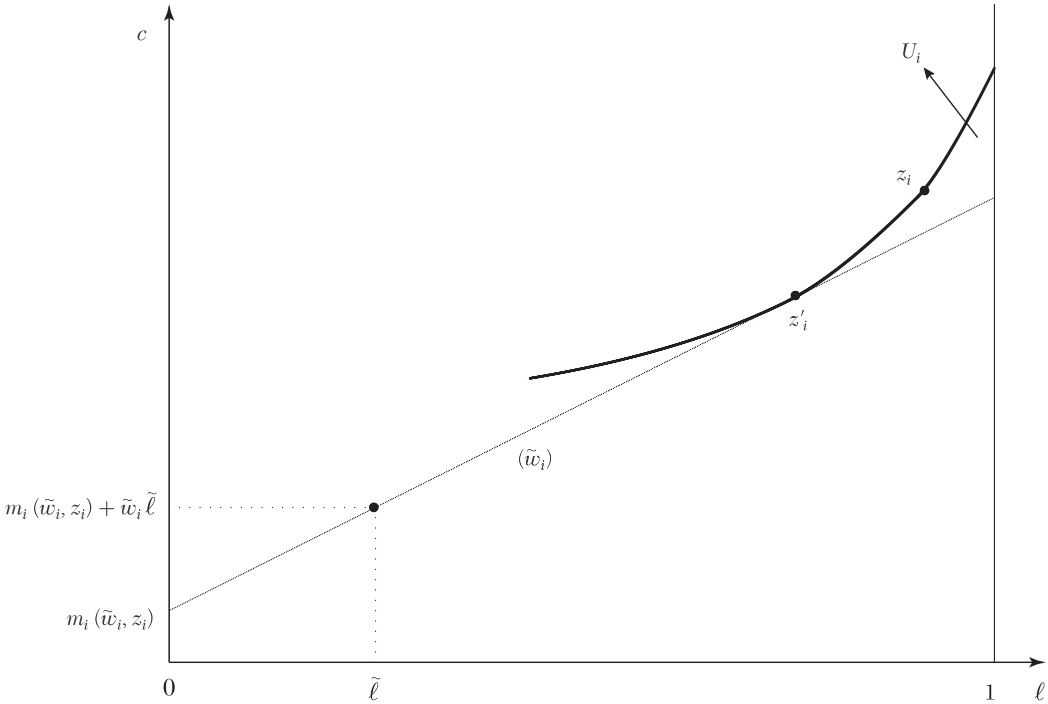

Figure 3. The Construction of Well-Being Indices

gives clear priority to the compensation goal of eliminating inequalities due to skills. In contrast, as explained in section 5, it applies the pairwise laissez-faire objective (i.e., seek equal lump-sum transfers) to individuals with equal skills only when their wage coincides with $\tilde{w}$. The laissez-faire objective is then less prominent.

In contrast, the latter index produces parallel budget lines for individuals with the same wage, and therefore seeks to make the implicit budgets equal for such individuals. When redistribution is made by lump-sum transfers in the first-best context, it then dutifully gives the same lump-sum transfers to individuals with the same wage. In the second-best context, it seeks to obtain a similar pattern for the implicit budgets. But inequalities due to unequal skills are less of a priority for this index. Because individuals with the same preferences may have implicit budgets that cross (when their wages differ), their ranking according to the $m_i(w_i, z_i) + w_i\tilde{\ell}$ index may not always coincide with the ranking of their indifference curves. This is illustrated in figure 4. Coincidence is guaranteed only for individuals who have a strong preference (of an almost Leontief sort) for working exactly $\tilde{\ell}$ hours (or in the generalization using a reference preference relation, only for individuals with personal preferences identical to the reference preferences).

This shows that the choice between the egalitarian-equivalent $m_i(\tilde{w}, z_i)$ and the libertarian $m_i(w_i, z_i) + w_i\tilde{\ell}$ is a choice

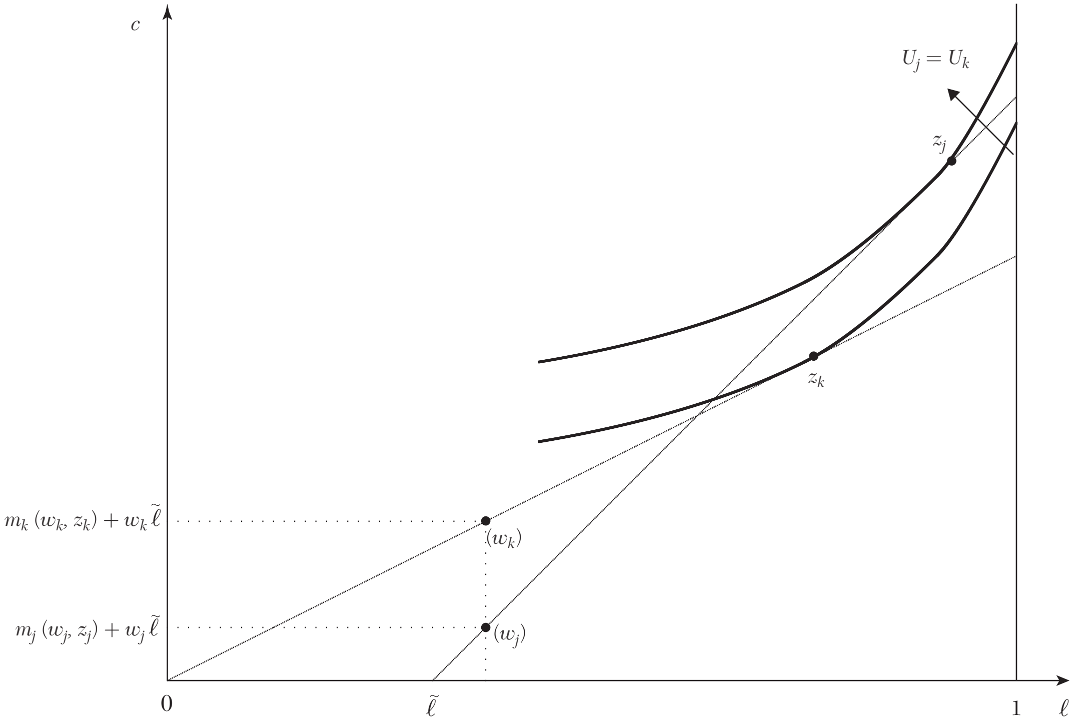

Figure 4. $U_j(z_j) > U_k(z_k)$ whereas $m_k(w_k,z_k) + w_k\tilde{\ell} > m_j(w_j,z_j) + w_j\tilde{\ell}$

between compensating unequal skills and laissez-faire among individuals with identical skills: pairwise compensation versus pairwise laissez-faire. And, as explained in the previous question, when $\tilde{\ell} \to 0$, the $m_i(w_i,z_i) + w_i\tilde{\ell}$ leans toward laissez-faire *tout court*.

Remember that in section 5.1 we showed that the libertarian approach, in the presence of public goods g, can be applied with money-metric utilities that incorporate a reference value $\tilde{g}$ of the public goods. The pure libertarian choice $\tilde{w}_i = w_i$ and $\tilde{\ell} = 0$ is then compatible with some redistributive income taxation if there is a correlation between $m_i(w_i,z_i)$, with $z_i = (\ell_i,c_i,g)$, and y_i. In the special case in which public goods affect productivity rather than utility, one works with $m_i(w_i(\tilde{g}),z_i)$. It is then possible to be in a situation in which $w_i(\tilde{g}) = (1-\lambda)\,w_i$; for instance, if $w_i(g) = w_i h(g)$ and $\tilde{g} < g$. Therefore, one sees that the introduction of the intermediate and personalized reference wage $\tilde{w}_i = (1-\lambda)\,w_i$ may also be useful to capture the virtual wage the individual would have under the reference quantity of public goods.

8.1.4 *Does One Want to Pay Special Attention to the Individuals with High or Low Aversion to Work?*

When one gives priority to inequalities in skills and wages, one cannot apply the laissez-faire principle to a great extent. One virtue of the laissez-faire principle is that it is neutral with respect to individual preferences. In a group of agents with identical skills, it seeks to give them the same implicit budget, disregarding their preferences.

Such neutrality is necessarily lost, then, when one focuses on eliminating inequalities due to skills, i.e., on the compensation objective. That this is logically necessary has been well established in the literature. It is due to the fact that one cannot at the same time give the same budget to individuals having identical wages and give bundles on the same indifference curves to individuals with identical preferences.[44]

When neutrality is lost, one has to decide what preferences to favor. This is where the choice of the parameter $\tilde{w}$ in $m_i(\tilde{w}, z_i)$ plays a role. With a low value for $\tilde{w}$, individuals who are more averse to work tend to obtain lower implicit budgets than individuals who are less averse. And the contrary occurs with a high value for $\tilde{w}$. This is illustrated in figure 5. Therefore, with a social-ordering function displaying a strong degree of inequality aversion, the work-averse individuals are better treated under a low $\tilde{w}$ than under a high $\tilde{w}$. If one considers, for instance, that work aversion in this model may be partly due to low job quality for the unskilled, a feature that is not well captured by this simple model, it may be prudent, or charitable, to choose a low value for $\tilde{w}$.

Two additional considerations suggest that the choice of $\tilde{w} = w_{\min}$ is worth considering seriously. First, it is endogenous to the wage distribution and implies that $\tilde{w}$ is the common wage when all individuals have the same wage, which itself entails that the laissez-faire is an optimal allocation in this case. Second, and more specifically, it is the only value in the $[w_{\min}, w_{\max}]$ interval that guarantees that redistribution will never violate the participation constraint (this constraint stipulates that every i should never prefer the $(0, 0)$ bundle to his assigned labor–consumption bundle).[45] One could, of course, add a participation constraint to the search for the optimal tax, but it seems preferable to make sure that the social objective itself guarantees that it will be satisfied by the optimal redistribution, whether in the first- or in the second-best context.[46]

8.2 *Discussion*

The previous subsection has shown that, even restricting attention to the class of indexes of the form

$$m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell},$$

there is a large spectrum of possibilities. This calls for two remarks. First, only a small subset of all the possible utility representations of given preferences are justified from a normative point of view. The properties discussed in this section provide a guide to the relevant ethical choices.

Second, one may think that all second-best efficient allocations may turn out to be justified by some appropriate choice of $\tilde{w}_i$ and $\tilde{\ell}$. This is definitely wrong, because at the laissez-faire allocation, the worst off for any of the social orderings in the class considered here include agents with the lowest wage $w_{\min}$. Therefore, an allocation that penalizes all of the unskilled, compared to the laissez-faire, can never be socially optimal for any of these orderings. We conjecture that many other restrictions could be found.

We close this section with a remark about a general impossibility to write the indexes discussed here as functions of a unique parameter gathering the preferences and skill heterogeneities. Let us

[44] For details, see Fleurbaey and Maniquet (1996, 2005).

[45] See Fleurbaey and Maniquet (2011a).

[46] See Fleurbaey (2008) and Fleurbaey and Maniquet (2011a) for further discussion of the choice of $\tilde{w} = w_{\min}$.

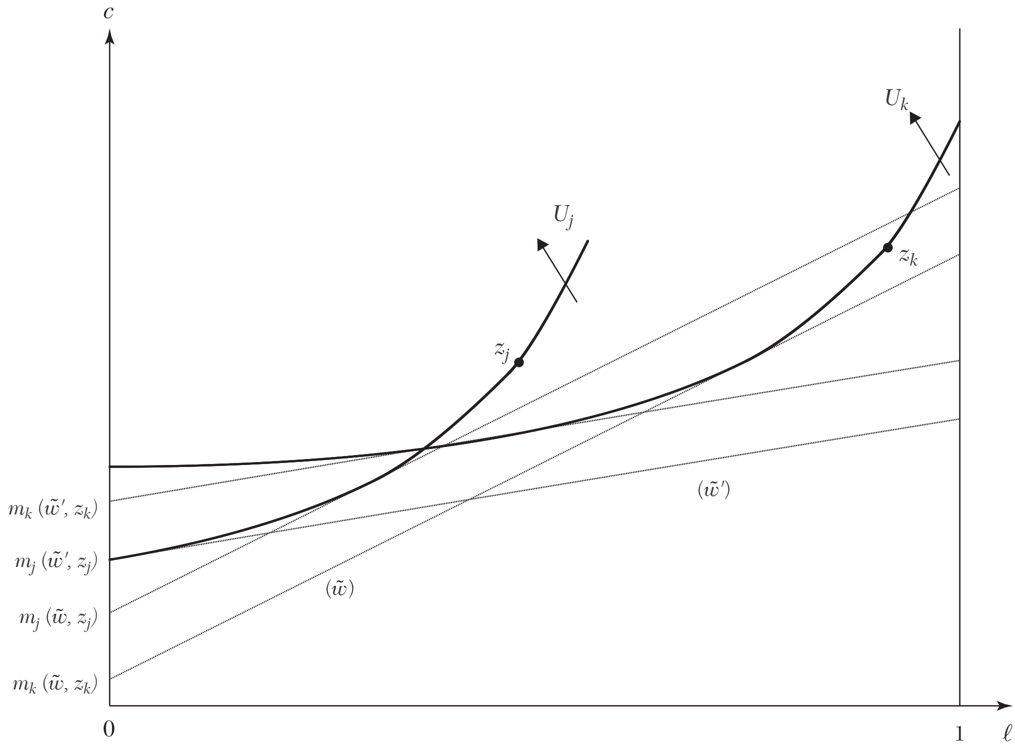

Figure 5. $m_j(\tilde{w},z_j) > m_k(\tilde{w},z_k)$ whereas $m_k(\tilde{w}',z_k) > m_j(\tilde{w}',z_j)$

consider, for instance, the reference-wage egalitarian-equivalent social ordering. Let us assume that preferences are quasi-linear and can be represented by the following utility function: $U_i(\ell_i, c_i) = c_i - v(\ell_i/\theta_i)$, for some increasing and convex v function satisfying $v(0) = 0$. With this utility function, one has $V_i(y_i, c_i) = c_i - v(y_i/(w_i\theta_i))$, so that the parameter $n_i = w_i\theta_i$ completely captures the behavioral heterogeneity of the population.

Let $\ell^*(\theta_i)$ be the optimal labor time of agent i if her wage were the reference $\tilde{w}$. The quasi-linearity assumption precisely guarantees that it is fixed and only depends on θ_i (once the common v is given). Then one computes

$$m_i(\tilde{w}, z_i) = c_i - v\left(\frac{y_i}{n_i}\right) - \tilde{w}\,\ell^*(\theta_i) + v\left(\frac{\ell^*(\theta_i)}{\theta_i}\right).$$

It is transparent that this utility index cannot be written as a function of the n_i parameter only. This implies that optimal tax analysis cannot then be performed with a unidimensional screening method, in general.

Here, one may worry that the class of utility indexes described in this section complicates, rather than enhances, the work of tax analysts. The next section examines what can be done.

9. *Applications of Taxation Theory*

In the previous section, we have spanned a wide set of possible social objectives from utilitarianism to libertarianism and introduced a subclass of intermediate objectives that involve money-metric utilities. As already explained in the previous sections, defining the social objective is only the beginning of the analysis, and deriving policy conclusions requires the additional work that makes the bulk of tax theory. In this section, we briefly review methodologies and provide illustrations.

It is useful to distinguish two types of applications. First, when a small set of tax options is offered, e.g., when a few possible reforms are considered, how can they be ranked in order to select the best according to the chosen social objective? This is the typical context of practical policy decisions. The political process is always full of political constraints that limit the set of options that can be considered. The second type of application, which has attracted much more attention in economic theory, is the "utopian" context in which the whole set of feasible taxes is considered and one can pick the best. It is important to see that the two contexts require different tools. Knowing the optimal tax formula does not provide a recipe for comparing two suboptimal taxes, and conversely, a simple criterion for the comparison of arbitrary taxes does not provide the optimal formula. It is, in particular, incorrect to draw lessons about reforms from the qualitative shape of the optimal tax. For instance, if one finds that the optimal tax is U shaped in marginal tax rates (as in Diamond 1998), this does not mean that any reform moving the tax formula toward a U shape is an improvement.

9.1 *Assessing Feasible Taxes*

The literature has mainly focused on the so-called "reform" problem (introduced by Feldstein 1976), which is the identification of marginal changes that improve the tax. The (weighted or unweighted) utilitarian social welfare function lends itself well to this analysis, as shown in formula (2), which displays weights for each local marginal change $\delta T(y)$ that simply aggregate the marginal social value of money for people earning y, and can be derived from knowledge of the statistical characteristics of the population. The Saez–Stantcheva approach proposes to derive these weights directly from intuitions about how deserving the population earning such or such level of income is. A systematic study of the reform problem, focusing on linear taxes, can be found in Guesnerie (1995).

Kaplow (2008) proposed to decompose every policy project (not just a tax reform) into an efficiency component and a distributional component. The idea is that, under the separability assumptions that make the Atkinson and Stiglitz (1976) result valid,[47] one can always decompose a change in allocation due to a particular project into two steps: i) the change in the reform under consideration, accompanied by an extra (virtual) adjustment in the income tax that is budget neutral and produces a new distribution of welfare that is Pareto comparable to the initial distribution; and ii) removing the extra adjustment in the income tax, which yields the final outcome of the reform. The idea is that, if the intermediate allocation is Pareto superior to the initial allocation, there is a sort of Kaldor–Hicks argument in favor of the reform: the reform could produce a Pareto-improving change if accompanied by a suitable reshuffling of the income tax. Of course, the distributional

[47] Individual preferences about commodities must be identical and independent of the quantity of labor. See Laroque (2005) for details.

impact of the second step should not be forgotten, as it may be substantial in some reforms. In a similar fashion to the Kaldor–Hicks approach, this type of analysis raises two possibilities: either the compensation is made and the Pareto principle is sufficient to evaluate the reform, or the compensation is not made, and potential Pareto superiority is not a decisive argument in favor of the reform.

The maximin criterion generally simplifies the analysis of reforms by reducing the number of income levels with a positive weight to a small set, possibly a singleton. In this case, a marginal reform is an improvement if it increases income at this particular level.

The combination of the maximin criterion with the money-metric utilities $m_i(\tilde{w}_i, z_i) + \tilde{w}_i\tilde{\ell}$ discussed in the previous section produces an extra simplification that makes it possible to evaluate all reforms, marginal and non-marginal, in a way that requires very little information about the population characteristics. This is a direct generalization of the analysis made in section 7 for the special case of $\lambda = 1$ and $\tilde{w} = w_{\min}$. Indeed, imagine an allocation (which may or may not be feasible by a tax) for which the individual money-metric utilities are all equal to a given k^*. Focus on the group of individuals sharing the same w_i. Equality in their utilities means that their implicit budget (i.e., the budget with slope $\tilde{w}_i$ that is tangent to their indifference curve in (ℓ, c) coordinates) is, for all of them, the same set

$$\left\{(\ell, c) \in X \mid c = k^* + \tilde{w}_i\left(\ell - \tilde{\ell}\right)\right\}.$$

In (y, c) coordinates, this corresponds to the set

$$B(k^*, w_i) = \left\{ (y, c) \in [0, w_i] \times \mathbb{R}_+ \mid c = k^* + \tilde{w}_i\left(\frac{y}{w_i} - \tilde{\ell}\right)\right\}.$$

Incidentally, note that, for any arbitrary allocation, one can compute the money-metric utility of an individual i in the (y, c) space by computing k^* such that $B(k^*, w_i)$ is tangent to i's indifference curve.

Take the union of all the sets $B(k^*, w)$, for $w \in [w_{\min}, w_{\max}]$, and construct the function $f(y)$ that espouses the upper bound of this set. It is a matter of simple algebra to compute that

$$f(y) = \begin{cases} k^* + \left(\lambda\tilde{w} + (1-\lambda)w_{\min}\right)\left(\frac{y}{w_{\min}} - \tilde{\ell}\right) & \text{for } y \leq w_{\min} \\ k^* + \left(\lambda\tilde{w} + (1-\lambda)y\right)\left(1 - \tilde{\ell}\right) & \text{for } y \geq w_{\min} \end{cases}.$$

The graph of this function is a very simple piecewise linear curve, with slope $\lambda\tilde{w}/w_{\min} + (1-\lambda) \geq 0$ for $y \leq w_{\min}$, and $(1-\lambda)\left(1-\tilde{\ell}\right) \in [0, 1]$ for $y \geq w_{\min}$.[48] Note that f is always concave, since $(1-\lambda)\left(1-\tilde{\ell}\right) \leq \lambda\tilde{w}/w_{\min} + (1-\lambda)$.

The graph of the function f provides a very simple graphical tool to evaluate taxes. Consider any given tax T and its induced allocation, and compute k^* so that the graph of f is tangent from below to the graph of $y - T(y)$. By construction, there is a w such that the curve $y - T(y)$ touches, without crossing, the set $B(k^*, w)$ from above for a level y_0. Assume that among the group of individuals sharing this particular w, there is one, i_0, whose income is y_0. The money-metric utility of i_0 is necessarily k^*—it cannot be greater, since i_0's bundle belongs to $B(k^*, w)$, and it cannot be lower, since i_0's indifference curve is weakly above $y - T(y)$, therefore weakly above $B(k^*, w)$. Moreover, this agent is necessarily among the worst off of the w-group, because the indifference curves of everyone in this group are weakly above $y - T(y)$. *A fortiori*, for the agents whose $B(k^*, w)$ is strictly below the

[48] When the set of values of w is discrete, the function is slightly less simple above $w_{\min}$. The line of slope $(1-\lambda)\left(1-\tilde{\ell}\right)$ is replaced by a succession of lines of slope 0 and 1, with the points $f(w)$ being on the line of slope $(1-\lambda)\left(1-\tilde{\ell}\right)$. For details, see Fleurbaey and Maniquet (2011c).

graph of $y - T(y)$, their money-metric utility must be above k^*. This means that i_0 is among the worst off of the whole population, and that k^* is the lowest utility.

If we assume that it is a good approximation to consider that every group of agents with a given w is spread everywhere on the budget curve $y - T(y)$ over the range $y \in [0, w]$, one sees that computing k^* by seeking tangency between the graph of f and the graph of $y - T(y)$ provides a good approximation of the lowest utility. Therefore, when k^*, thus computed, is greater for one tax than for another, social welfare is greater with the corresponding tax.

The computation of k^* requires no knowledge at all of the characteristics of the population, except the value of $w_{\min}$ and the assumption that there is sufficient diversity of preferences so that individuals within each wage group are spread on their common budget curve.

Observe that the concavity of f means that a basic form of progressivity is built in to the analysis. Of course, an optimal tax may still have declining marginal tax rates if efficiency requires it, due to incentives.

A few salient cases are worth describing. First, for $\lambda = 1$ (egalitarian-equivalent criterion) or $\tilde{\ell} = 1$ (focus on full incomes for the implicit budgets), the slope is nill after $w_{\min}$, which means that the comparison of taxes bears only on the graph of $y - T(y)$ below $w_{\min}$, reflecting an absolute priority granted to low incomes.

Second, for $\lambda = 0$ (priority to laissez-faire) or $\tilde{w} = w_{\min}$, the slope is one below $w_{\min}$, implying that the goal is to have a zero marginal tax on low incomes. The particular reference-wage egalitarian-equivalent criterion studied in section 7, with $\lambda = 1$ and $\tilde{w} = w_{\min}$, has a unit slope until $w_{\min}$, and zero beyond, meaning that the goal is to give a lump-sum transfer to low incomes, and to maximize it.

Third, the libertarian criterion, with $\lambda = 0$ and $\tilde{\ell} = 0$, has a unitary slope throughout and identifies as the worst off those who pay the most taxes. At the opposite extreme, the egalitarian-equivalent $(\lambda = 1)$ case with $\tilde{w} = 0$ has a zero slope throughout, implying a focus on the lowest consumption level and seeking to maximize the minimum income support. Note that for $\lambda > 0$, by choosing a high $\tilde{w}$ there is no upper bound to the slope before $w_{\min}$. In contrast, the slope above $w_{\min}$ is always between zero and one.

Let us briefly discuss the case of public goods affecting productivity. For this case, we saw in the previous section that the libertarian approach was compatible with taking $m_i(\tilde{w}_i, z_i)$ for $\tilde{w}_i = (1 - \lambda) w_i$ if $w_i(g) = w_i h(g)$. The function f is then quite simple: it has slope $(1 - \lambda)$ throughout. This pushes the optimal tax toward a proportional tax, which is reminiscent of Smith's (1776) intuition about the benefit tax, as recalled by Weinzierl (2014a).

This methodology can also be used, in the context of a particular reform, to find the subset of criteria (in the class of maximin criteria applied to utilities of the $m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell}$ form) that are compatible with this reform. Assuming that the reform is motivated by a consistent objective, one can then retrieve the revealed preferences of the policy maker. For instance, Fleurbaey (2008) examines the US 1996 welfare reform and finds that it is an improvement for a reference-wage egalitarian-equivalent criterion (i.e., $\lambda = 1$) only if $\tilde{w}$ is sufficiently high (meaning that hardworking agents are favored enough by the decision maker).

Figure 6 illustrates the assessment of taxes with the f function. The budget reflects the 2013 US income tax for a couple with two children.[49] For an egalitarian-equivalent

[49]The data come from the OECD Tax Benefit Calculator (http://www.oecd.org/els/benefits-and-wages-models.htm) based on the OECD report on benefits and

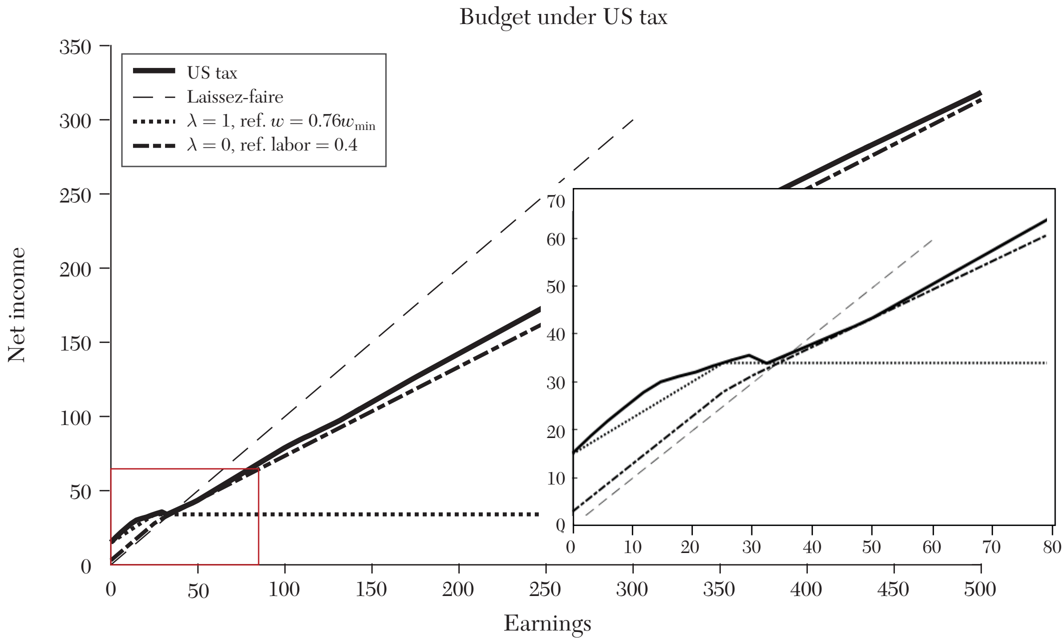

Figure 6. Evaluating the US Income Tax

approach ($\lambda = 1$) with a value of $\tilde{w}$ that is sufficiently low (at $0.76w_{\min}$), the earnings that receive positive weight and correspond to the lowest money-metric utility are zero and $w_{\min}$ (set here at \$25K). A lower value for $\tilde{w}$ would single out the zero earnings, a greater value would focus on the minimum wage (working poor).[50] For a libertarian criterion ($\lambda = 0$, with $\tilde{\ell} = 0.4$ on the figure), the focus is not just on the low incomes (around \$50K), but also the high incomes. In fact, since the marginal tax rate is slightly above 0.4 for high incomes, the function f shown on the figure actually crosses the budget curve around \$650K. Therefore, the worst off are actually the top incomes for this particular f. The value of $\tilde{\ell} = 0.439$ (corresponding to marginal tax rate for high incomes) is the threshold below which the top incomes get the priority and above which the priority shifts to lower incomes.

9.2 *Optimal Tax*

Let us very briefly discuss the first-best context in which incentive constraints do not apply. It was recalled in section 4 that utilitarianism penalizes the more productive individuals for given preferences. The maximin criterion naturally avoids this problem for utility indexes such that individuals with identical preferences are given the

wages (http://www.oecd.org/els/soc/benefits-and-wages-country-specific-information.htm) for the United States. We are grateful to Dirk Neumann for guiding us to these data.

[50] Note that the budget curve is not monotonic, due to the phasing out of social assistance, and that the trough at \$33K is also tangent to f. Our analysis is compatible with preferences failing to be monotonic in leisure, in which case some individuals could end up picking a point in the non-monotonic part of the curve.

same utility index. This property holds true with the money-metric utilities $m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell}$ if $\lambda = 1$. In contrast, when $\lambda < 1$, the utility index depends not just on the individual's preferences, but also on w_i. The first-best allocations equalize these utilities, which means that individual i is indifferent to a budget with slope $\tilde{w}_i$ and all such budgets give the same consumption for $\tilde{\ell}$. If $\tilde{\ell} > 0$, the budget lines cross and the lower part of these budgets (below $\tilde{\ell}$) are therefore less favorable for the individuals with greater w_i. It is then possible to see some pairs of individuals with identical preferences in a configuration such that the one with greater wage in the pair is on a lower indifference curve. This will happen for individuals who are sufficiently averse to work so that their optimal choice of labor in these implicit budgets are below $\tilde{\ell}$. When $\tilde{\ell} = 1$, which implies equalizing full incomes at the optimum, this becomes a general pattern, as with utilitarianism. This has famously been called the "slavery of the talented" by Dworkin (1981b).[51] In summary, with the $m_i(\tilde{w}_i, z_i) + \tilde{w}_i \tilde{\ell}$ class of utilities, if one wants to fully avoid the risk of penalizing any individual for being more productive, one has to pick either $\lambda = 1$ or $\tilde{\ell} = 0$.[52]

In the second-best context, the computation of the optimal tax formula, or at least the marginal tax rates, has been the focus of the literature since Mirrlees (1971). In practice, it is always possible to find the optimal tax with a sufficiently powerful computer for any given sample describing the population characteristics. However, understanding the relation between the characteristics of the population, the social objective, and the optimal tax is an important theoretical question, and provides valuable insights into the exact form of the information that is needed to compute the tax (sufficient statistics).

The typical tax formulas (as derived in Atkinson and Stiglitz 1980, Diamond 1998, Saez 2001, Jacquet and Lehmann 2014, and Saez and Stantcheva 2016) contain two endogenous terms on the right-hand side of the marginal tax equation: the average marginal social value of income above the level of earnings y, and the elasticity of earnings with respect to the net income rate (one minus the tax rate). For instance, the formula obtained by Jacquet and Lehmann is:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\epsilon(y)} \frac{1-H(y)}{yh(y)} \left(1 - \frac{\int_y^{+\infty} [g(z) + \eta(z)\, T'(z)] h(z)\, dz}{1-H(y)}\right), \tag{3}$$

where $\epsilon(y)$ is the average elasticity of y with respect to $1 - T'$ among agents earning y, h and H are the PDF and CDF of the distribution of y at the allocation, $g(z)$ is the marginal social value at earning level z, and $\eta(z)$ is the average derivative of y with respect to a lump-sum transfer among agents earning z. In order to implement such formulas, it is then common to assume a fixed distribution of marginal social values (and/or a fixed distribution of income) and a constant elasticity, and when possible an iteration can be performed.

The purpose of this section is to illustrate the various optimal taxes that are obtained for a population calibrated to resemble the US population, and with quasi-linear preferences represented by $c - a_i \ell^{1+1/\varepsilon}$. The calibration is similar to Lockwood and Weinzierl's (2015), with a few changes. First, a full-time work bound is introduced, which matters in the definition of money-metric utilities. Second, the diversity of preferences is specified since some social objectives require knowing the joint distribution of wages and preferences, even if behavioral heterogeneity is unidimensional, as

[51] The prospect that such a tax might force the talented to work in order to pay their tax liability is discussed in the law literature (see Zelenak 2006 and Hasen 2007).

[52] In the case of identical preferences over consumption and leisure, it is enough to pick $\tilde{\ell}$ below the range of labor values the individuals would choose in their implicit budgets.

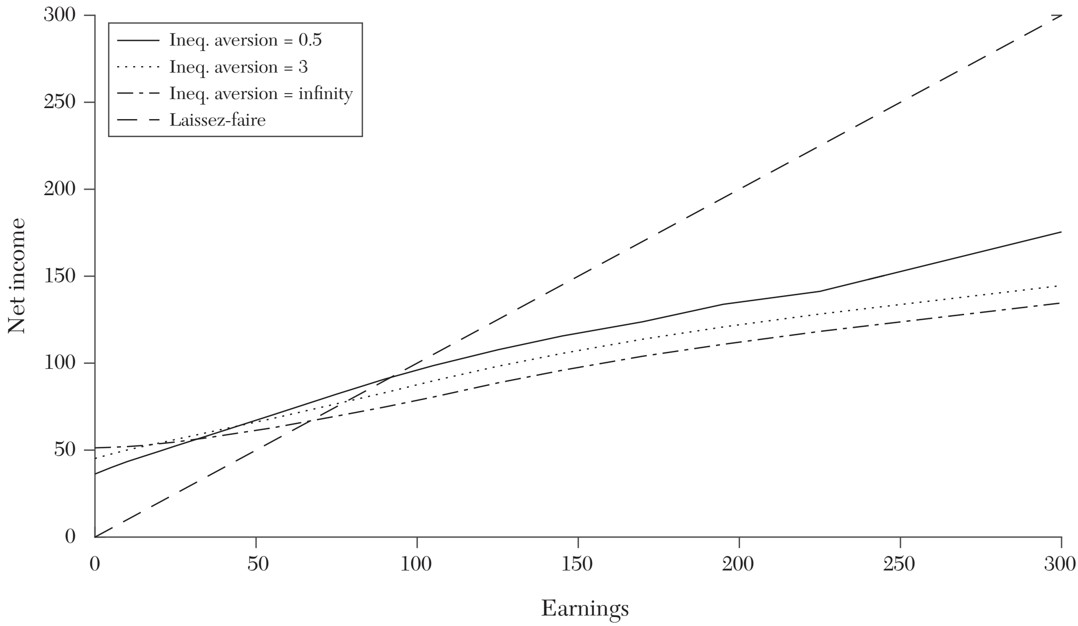

Figure 7. Utilitarian Taxes (for $\rho = 0.5, 3, +\infty$)

explained at the end of the previous section.[53] With this population, many agents work full time, so that the first-order approach underlying formula (3) cannot be applied. Moreover, the maximin money-metric criteria would provide weights that can vary abruptly with small changes in the allocation and therefore do not lend themselves to the use of this formula (except when the marginal social value is zero, but for the lowest income). We therefore focus on the computation of an optimal piecewise linear tax.[54]

Figure 7 displays the various taxes obtained with a utilitarian social welfare function $\frac{1}{1-\rho}\sum_i u_i^{1-\rho}$ for different values of ρ. The utilities adopted here are simply the intercept utilities $c - a_i \ell^{1+1/\varepsilon}$, which can be interpreted as the amount of consumption that i would accept in absence of work—

[53] The population contains 300 types of households with wages w_i distributed according to a lognormal distribution with parameters $(\mu, \sigma) = (2.2, 0.6)$ so that the distribution of earnings fits the quintiles of the distribution of pretax household income from the census for 2014 for a flat tax of 30 percent ($29k, 53k, 82k, 129k, and 230k for the 95th percentile). A household is considered in our application to be like 1.66 adults (census), so that for an individual minimum wage of $15,000 per year we use a household minimum of $25,000. The parameters a_i of the utility functions $c - a_i \ell^{1+1/\varepsilon}$ take three values so that the percentage of households working full time (more than 90 percent), part time, and less than 1/3, at the status quo, is 45 percent, 43 percent, and 12 percent. We have introduced a greater variety of preference parameters for the minimum-wage households so that their labor supply spreads over the whole interval. The parameter ε is set at 0.5 to match a reasonable estimate of the elasticity of part-time workers, and the average elasticity of the whole population at the status quo is 0.26.

[54] There are sixteen tax brackets. The maximum income in the sample is 500.

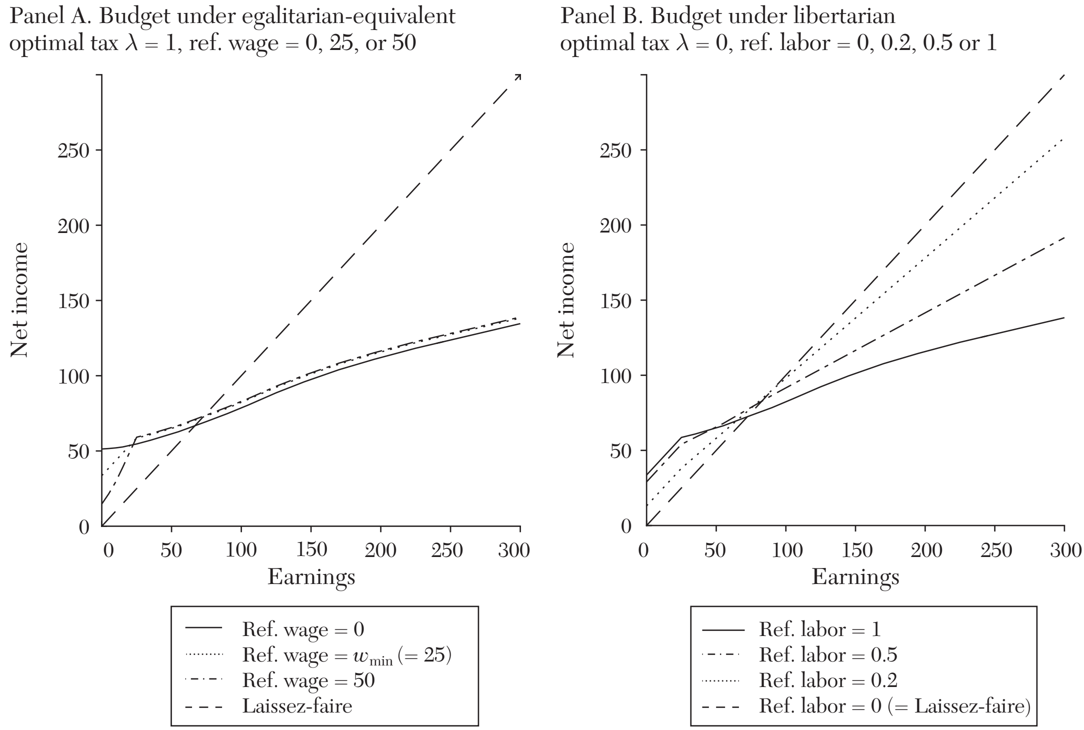

Figure 8. Optimal Egalitarian-Equivalent and Libertarian Taxes

these are the money-metric utilities for $\lambda = 1$ and $\tilde{w} = 0$. This choice of cardinal utilities is of course only one possibility among many. As one can see on the figure, the marginal tax rates go from an inverted U shape to a declining curve, and then to a U-shaped pattern (Diamond 1998) as inequality aversion increases.

Figure 8, on the left graph, shows various taxes obtained with the reference-wage egalitarian-equivalent criteria, i.e., when $\tilde{w}_i = \tilde{w}$ for all agents. Given the absence of income effects and the fact that only the situation of the income below $w_{\min}$ matters for these criteria, the budget curves are parallel for $y > w_{\min}$. Note that in this model with quasi-linear preferences, the criterion with $\tilde{w} = w_{\min}$ has the same optimal tax as Roemer's equal opportunity criterion that maximizes the average utility of the unskilled agents whose wage is $w_{\min}$ (assuming that intercept utilities are used in Roemer's criterion, as in the above simulations of utilitarianism).

Figure 8, in the right graph, also shows various taxes obtained with the libertarian criteria, i.e., when $\tilde{w}_i = w_i$ and $\tilde{\ell}$ varies from 0 (pure libertarianism) to 1 (which is the most redistributive in this category and coincides with the most redistributive utilitarian tax in figure 7).[55] Note that the US tax

[55] For these criteria, the function f introduced in the previous subsection has slope 1 until $w_{\min}$ and $1 - \tilde{\ell}$ beyond. The optimal tax for $\tilde{\ell} = 0.2$ produces a budget that actually espouses the graph of f for a well-chosen k^*.

shown on figure 6 bears some resemblance with the libertarian tax one would obtain with $\tilde{\ell} = 0.4$, since it has a marginal tax close to zero for low incomes, and then close to 40 percent (although around 30 percent for certain intermediate levels of earnings).

For a value of λ between 0 and 1, one obtains optimal taxes that are intermediate between those of the two sides of figure 8. For $\lambda = 1/2$, with $\tilde{w} = 0$ and $\tilde{\ell} = 0$ (a case in which f has constant slope 1/2, and may arise for a libertarian approach taxing benefits from public goods, considering that half of everyone's wage can be attributed to the benefits brought by public goods), in our calibrated economy one obtains an optimal tax that makes $y - T(y)$ coincide with $f(y)$ and is a flat tax (with a marginal tax rate equal to 50 percent throughout).

10. *Conclusion*

In this paper, we have examined the contribution that notions of fairness can make to optimal income taxation theory. Recent interest in fairness principles capturing the relevant differences between deserved and undeserved income, as well as between circumstances and effort, the importance of laissez-faire, and the problems with tagging, makes it timely to connect public economics with the theory of fair allocation, which provides useful concepts.

While some authors have argued for a radical overhaul of taxation theory that would throw out the welfare economics baby with the utilitarian bath water, we have pleaded for going beyond the conventional utilitarian criterion while retaining the social welfare function and its arguments, the utility functions. Specifically, we have shown that the individual utility indexes are malleable tools that can incorporate many of the fairness considerations listed in the previous paragraph. Perhaps the current weariness with the social welfare function comes from an exclusive focus on weighted variants of the utilitarian criterion. The utility-weighting approach, indeed, is limited because the weights cannot be transparently connected to fairness ideas. Even the method of weighting incomes directly, as proposed by Saez and Stantcheva (2016), cannot always be applied easily because identifying the levels of income that deserve a positive weight may sometimes require a detour by a suitably defined social welfare function involving appropriate utility indexes.

We have focused on the maximin criterion in a large part of the paper. As we have stressed, however, less extreme degrees of inequality aversion than the maximin can be studied, with the same set of possible utility functions as listed in section 8. The maximin is worth considering for two reasons. First, it is important to recognize that a distinction must be made between redistribution and inequality aversion in the social welfare function. The maximin is compatible with very little redistribution, even with the full libertarian approach, since a crucial element in deciding how much to redistribute is the measure of individual advantage (the utility function), which can take account of endowments, desert, and similar considerations. The second reason for paying attention to the maximin is that it makes the comparative evaluation of taxes easier, as shown in section 9.1, where a very simple graphical tool has been provided, which can be adapted to the whole array of utility functions presented in section 8.

Two approaches have been highlighted, with all intermediate cases being possible. The egalitarian-equivalent approach focuses on low incomes, and lets the decision maker choose to prioritize the poor workers who

More generally, when the graph of f corresponds to an efficient budget, it is associated to the optimal tax (Fleurbaey and Maniquet 2011c).

work full time, or those who work less (or not at all). The libertarian approach is neutral among poor workers and, in its moderate version, can accept a certain degree of redistribution represented by a graphical rod featuring a flat tax above low incomes. Unlike the egalitarian-equivalent approach, it does pay attention to all incomes, and can declare the worst off to be among the top incomes, as in the example provided for the US tax in figure 6.

We hope that the last two sections will help optimal tax theorists and applied analysts to incorporate the relevant fairness considerations of their choice easily, via a suitable choice of the utility indexes.

The analysis in this paper has been focused on the standard income tax model, and similar considerations can be applied to related standard models of taxation or new models incorporating additional features. It appears that both in theory and in popular discourse, different contexts often call for different fairness principles. For instance, inheritance taxation raises issues about altruism toward descendants that do not appear in the Mirrlees model. The methodology laid out in this paper is, in principle, applicable to a wide set of taxation problems, such as capital income taxation, commodity taxation, or Pigovian taxation.[56]

Such extensions can also serve to address the worry that the preferences over labor and consumption that play an important role in the Mirrlees model and in usual applications of optimal tax theory may be influenced by factors for which the laissez-faire principle is not justified. Some workers may be more averse to work than others because they only have access to less pleasant or more dangerous jobs, or because they have children or relatives needing their care at home, or because their health reduces their ability to do certain tasks. As noted in section 8, the worry that greater work aversion may be explained by disadvantages can partly be addressed by answering the fourth question in the previous section in a particular way, by selecting a low reference wage in the construction of the utility index. However, addressing these issues completely and satisfactorily requires adding the relevant features into the model, and, for applications, finding estimates of the distribution of characteristics in the relevant population.

An important extension of the Mirrlees model that has been developed recently involves dynamic processes (see Golosov, Tsyvinski, and Werning 2006 for a detailed introduction to this literature). In the dynamic optimal taxation literature, the focus is on the ability of the tax system to allow the economy to reach constrained efficient allocations in the presence of individual and/or aggregate uncertainty about future productivity (see, e.g., Farhi and Werning 2013) or government expenditures (see, e.g., Kocherlakota 2005); the impossibility of the government to commit not to use the information revealed by agents' past choices (see, e.g., Berliant and Ledyard 2014); and the advantage of using age or history-dependent taxation schemes (see, e.g., Weinzerl 2011).

To the extent that they deal with efficiency, the results obtained in that literature are of interest to all approaches that are consistent with efficiency, which includes libertarian and resource-egalitarian approaches. The potential contribution of fairness approaches to dynamic optimal taxation questions should be, as in the static optimal taxation setting examined in this paper, to help design the individual utility indices that could be used when the goal of the research is to go beyond efficiency analysis and to study dynamic redistribution.

[56] See, e.g., Ambec and Ehlers (2016) on externalities and Fleurbaey (2006) on commodity taxes. The study of inheritance taxation in Piketty and Saez (2013a) relies on the maximin criterion and could be extended to heterogeneous utilities with the help of money-metric utilities. One also finds a brief analysis of capital income taxation with money-metric utilities in Fleurbaey (2008).

In the framework of the resource-egalitarian approach, for instance, specific ethical questions include: how to compare agents with different rates of time preference, how to treat agents facing different productivity paths, how to treat agents facing different productivity shocks, and so on.

Note that time is, conceptually, not hard to incorporate in the libertarian and resource-egalitarian approaches, since it only corresponds to increasing the number of commodities. In contrast, insofar as risk is an essential ingredient in dynamic problems, the complications mentioned at the end of section 4 become relevant and, outside utilitarianism, the differences between ex ante criteria that bear on individuals' expected utilities or certainty-equivalent consumptions, and ex post criteria that rely on the expected value of social welfare, have strong implications for policy conclusions.[57]

The behavioral literature raises an important challenge for all the approaches that rely on individual preferences, i.e., all the approaches reviewed in this paper. Individuals may not be aware of their budget options (Chetty, Looney, and Kroft 2009; and Chetty, Friedman, and Saez 2013), and their preferences can be unstable and unreliable (see the review in Shafir 2016). The adaptation of welfare criteria to behavioral complications is an important topic that lies beyond the scope of this paper and remains quite open.[58]

[57] For instance, an application of resource-egalitarian criteria with money-metric utilities to the problem of redistribution between individuals facing a risk of early death, has been done by Fleurbaey, Leroux, and Ponthiere (2014) and Fleurbaey et al. (2016), and can be compared to the utilitarian study by Bommier, Leroux, and Lozachmeur (2011). It is shown that, while an ex ante approach only seeks to redistribute between groups with different life expectancies, an ex post approach, in the second best, seeks to encourage a declining profile of consumption over the life cycle in order to reduce the disadvantage suffered by the individuals who die prematurely.

[58] See the special issue of *Social Choice and Welfare* (vol. 38, no 4, 2012), and in particular the introduction by Sugden and McQuillin (2012). Relying on Bernheim and Rangel (2009), Fleurbaey and Schokkaert (2013) propose an adaptation of fairness criteria when individuals have a stable core of incomplete preferences.

References

Adler, Matthew D. 2012. *Well-Being and Fair Distribution: Beyond Cost–Benefit Analysis*. Oxford and New York: Oxford University Press.

Akerlof, George A. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *American Economic Review* 68 (1): 8–19.

Alesina, Alberto, and George-Marios Angeletos. 2005. "Fairness and Redistribution." *American Economic Review* 95 (4): 960–80.

Ambec, Stephan, and Lars Ehlers. 2016. "Regulation via the Polluter-Pays Principle." *Economic Journal* 126 (593): 884–906.

Arneson, Richard J. 1989. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56 (1): 77–93.

Arrow, Kenneth J. 1973. "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice." *Journal of Philosophy* 70 (9): 245–63.

Atkinson, Anthony B. 1995. *Public Economics in Action: The Basic Income/Flat Tax Proposal*. Oxford and New York: Oxford University Press.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6 (1–2): 55–75.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1980. *Lectures on Public Economics*. New York: McGraw Hill.

Bergson, Abram. 1938. "A Reformulation of Certain Aspects of Welfare Economics." *Quarterly Journal of Economics* 52 (2): 310–34.

Berliant, Marcus, and John O. Ledyard. 2014. "Optimal Dynamic Nonlinear Income Taxes with No Commitment." *Journal of Public Economic Theory* 16 (2): 196–221.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

Besley, Timothy, and Stephen Coate. 1995. "The Design of Income Maintenance Programmes." *Review of Economic Studies* 62 (2): 187–221.

Bierbrauer, Felix. 2009. "Optimal Income Taxation and Public Good Provision with Endogenous Interest Groups." *Journal of Public Economic Theory* 11 (2): 311–42.

Blackorby, Charles, and David Donaldson. 1988. "Money Metric Utility: A Harmless Normalization?" *Journal of Economic Theory* 46 (1): 120–29.

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

Boadway, Robin, and Michael Keen. 1993. "Public

Goods, Self-Selection and Optimal Income Taxation." *International Economic Review* 34 (3): 463–78.
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): 475–98.
Bommier, Antoine, Marie-Louise Leroux, and Jean-Marie Lozachmeur. 2011. "On the Public Economics of Annuities with Differential Mortality." *Journal of Public Economics* 95 (7–8): 612–23.
Bourguignon, François, and Amedeo Spadaro. 2012. "Tax-Benefit Revealed Social Preferences." *Journal of Economic Inequality* 10 (1): 75–108.
Brett, Craig, and John A. Weymark. 2003. "Financing Education Using Optimal Redistributive Taxation." *Journal of Public Economics* 87 (11): 2549–69.
Broome John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.
Cappelen, Alexander W., Astri Drange Hole, Erik Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–27.
Cappelen, Alexander W., Erik Sørensen, and Bertil Tungodden. 2010. "Responsibility for What? Fairness and Individual Responsibility." *European Economic Review* 54 (3): 429–41.
Chetty, Raj, John Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–721.
Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.
Chiappori, Pierre-André. 2016. "Welfare and the Household." In *Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, 821–43. Oxford and New York: Oxford University Press.
Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2003. "Optimal Defaults." *American Economic Review* 93 (2): 180–85.
Choné, Philippe, and Guy Laroque. 2010. "Negative Marginal Tax Rates and Heterogeneity." *American Economic Review* 100 (5): 2532–47.
Christiansen, Vidar. 1981. "Evaluation of Public Projects under Optimal Taxation." *Review of Economic Studies* 48 (3): 447–57.
Clark, Andrew E. 2016. "SWB as a Measure of Individual Well-Being." In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, 518–52. Oxford and New York: Oxford University Press.
Clark, Andrew E., Paul Frijters, and Michael A. Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature* 46 (1): 95–144.
Cohen, Gerald A. 1989. "On the Currency of Egalitarian Justice." Ethics 99 (4): 906–44.
d'Aspremont, Claude, and Louis Gevers. 2002. "Social Welfare Functionals and Interpersonal Comparability." In *Handbook of Social Choice and Welfare, Volume 1*, edited by Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura, 459–541. Amsterdam and San Diego: Elsevier.
Deaton, Angus. 1980. "Measurement of Welfare: Theory and Practical Guidelines." Living Standards Measurement Study Working Paper 7.
Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge and New York: Cambridge University Press.
Decancq, Koen, Marc Fleurbaey, and Erik Schokkaert. 2015. "Happiness, Equivalent Incomes and Respect for Individual Preferences." *Economica* 82 (s1): 1082–106.
Decoster Andre, and Peter Haan. 2015. "Empirical Welfare Analysis with Preference Heterogeneity." *International Tax and Public Finance* 22 (2): 224–51.
Dhillon Amrita, and Jean-François Mertens. 1999. "Relative Utilitarianism." *Econometrica* 67 (3): 471–98.
Diamond, Peter A. 1967. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: Comment." *Journal of Political Economy* 75 (5): 765–66.
Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1): 83–95.
Diener, Ed, John F. Helliwell, and Daniel Kahneman, eds. 2010. *International Differences in Well-Being*. Oxford and New York: Oxford University Press.
Di Tella, Rafael and Robert McCulloch. 2006. "Some Uses of Happiness Data in Economics." *Journal of Economic Perspectives* 20 (1): 25–46.
Dolan, Paul. 2014. *Happiness by Design: Change What You Do, Not How You Think*. New York: Penguin.
Dolan, Paul, and Mathew P. White. 2007. "How Can Measures of Subjective Well-Being Be Used to Inform Public Policy?" *Perspectives on Psychological Science* 2 (1): 71–85.
Dworkin, Ronald. 1981a. "What is Equality? Part 1: Equality of Welfare." *Philosophy and Public Affairs* 10 (3): 185–246.
Dworkin, Ronald. 1981b. "What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10 (4): 283–345.
Dworkin, Ronald. 2000. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge and London: Harvard University Press.
Farhi, Emmanuel, and Iván Werning. 2013. "Insurance and Taxation over the Life Cycle." *Review of Economic Studies* 80 (2): 596–635.
Feldstein, Martin. 1976. "On the Theory of Tax Reform." *Journal of Public Economics* 6 (1–2): 77–104.
Feldstein, Martin. 2012. "The Mirrlees Review." *Journal of Economic Literature* 50 (3): 781–90.
Fleurbaey, Marc. 2006. "Is Commodity Taxation Unfair?" *Journal of Public Economics* 90 (10–11):

1765–87.

Fleurbaey, Marc. 2008. *Fairness, Responsibility, and Welfare*. Oxford and New York: Oxford University Press.

Fleurbaey, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford and New York: Oxford University Press.

Fleurbaey, Marc, Marie-Louise Leroux, Pierre Pestieau, and Gregory Ponthiere. 2016. "Fair Retirement under Risky Lifetime." *International Economic Review* 57 (1): 177–210.

Fleurbaey, Marc, Marie-Louise Leroux, and Gregory Ponthiere. 2014. "Compensating the Dead." *Journal of Mathematical Economics* 51: 28-41.

Fleurbaey, Marc, and François Maniquet. 1996. "Fair Allocation with Unequal Production Skills: The No Envy Approach to Compensation." *Mathematical Social Sciences* 32 (1): 71–93.

Fleurbaey, Marc, and François Maniquet. 2005. "Fair Social Orderings When Agents Have Unequal Production Skills." *Social Choice and Welfare* 24 (1): 93–127.

Fleurbaey, Marc, and François Maniquet. 2006. "Fair Income Tax." *Review of Economic Studies* 73 (1): 55–83.

Fleurbaey, Marc, and François Maniquet. 2007. "Help the Low Skilled or Let the Hardworking Thrive? A Study of Fairness in Optimal Income Taxation." *Journal of Public Economic Theory* 9 (3): 467–500.

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

Fleurbaey, Marc and François Maniquet. 2011b. "Compensation and Responsibility." In *Handbook of Social Choice and Welfare, Volume 2*, edited by Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura, 507–604. Amsterdam: Elsevier, North-Holland.

Fleurbaey, Marc and François Maniquet. 2011c. "Kolm's Tax, Tax Credit, and the Flat Tax," In *Social Ethics and Normative Economics: Essays in Honour of Serge-Christophe Kolm*, edited by Marc Fleurbaey, Maurice Salles, and John A. Weymark, 217–39. Berlin: Springer.

Fleurbaey, Marc, and François Maniquet. 2012. *Equality of Opportunity: The Economics of Responsibility*. Hackensack and London: World Scientific.

Fleurbaey, Marc, and Erik Schokkaert. 2013. "Behavioral Welfare Economics and Redistribution." *American Economic Journal: Microeconomics* 5 (3): 180–205.

Fleurbaey, Marc, and Koichi Tadenuma. 2014. "Universal Social Orderings: An Integrated Theory of Policy Evaluation, Inter-society Comparisons, and Interpersonal Comparisons." *Review of Economic Studies* 81 (3): 1071–101.

Fleurbaey, Marc, Bertil Tungodden, and Howard F. Chang. 2003. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle: A Comment." *Journal of Political Economy* 111 (6): 1382–85.

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

Gahvari, Firouz. 2006. "On the Marginal Cost of Public Funds and the Optimal Provision of Public Goods." *Journal of Public Economics* 90 (6–7): 1251–62.

Golosov, Mikhail, Nayarana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies* 70 (3): 569–87.

Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning. 2007. "New Dynamic Public Finance: A User's Guide." In *NBER Macroeconomics Annual 2006: Volume 21*, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, 317–88. Cambridge and London: MIT Press.

Graham, Carol. 2009. *Happiness around the World: The Paradox of Happy Peasants and Miserable Millionaires*. Oxford and New York: Oxford University Press.

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

Hammond, Peter J. 1994. "Money Metric Measures of Individual and Social Welfare Allowing for Environmental Externalities." In *Models and Measurement of Welfare and Inequality*, edited by Wolfgang Eichhorn, 694–724. Berlin and Heidelberg: Springer.

Harsanyi, John C. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking." *Journal of Political Economy* 61 (5): 434–35.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–21.

Harsanyi, John C. 1976. Essays on Ethics, Social Behaviour, and Scientific Explanation. Dordrecht: D. Reidel.

Hasen, David. 2007. "Liberalism and Ability Taxation." *Texas Law Review* 85: 1057–113.

Hildreth, Clifford. 1953. "Alternative Conditions for Social Orderings." *Econometrica* 21 (1): 81–94.

Jacquet, Laurence, and Etienne Lehmann. 2014. "Optimal Nonlinear Income Taxation with Multidimensional Types: The Case with Heterogeneous Behavioral Responses." THEMA Working Paper 2014–01.

Jacquet, Laurence, and Dirk Van de Gaer. 2011. "A Comparison of Optimal Tax Policies When Compensation and Responsibility Matter." *Journal of Public Economics*, 95: 1248–62.

Kahneman, Daniel, Ed Diener, and Norbert Schwarz, eds.1999. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Kahneman, Daniel and Alan B. Krueger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24.

Kaneko, Mamoru. 1977. "The Ratio Equilibria and the Core of the Voting Game G(N, W) in a Public Goods Economy." *Econometrica* 45 (7): 1589–94.

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

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.

Kaplow, Louis, and Steven Shavell. 2002. *Fairness versus Welfare*. Cambridge and London: Harvard University Press.

Kocherlakota, Nayarana R. 2005. "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica* 73 (5): 1587–621.

Kolm, Serge-Christophe. 1999. *Justice and Equity*. Cambridge: MIT Press, 1972.

Kolm, Serge-Christophe. 2004. *Macrojustice: The Political Economy of Fairness*. Cambridge and New York: Cambridge University Press.

Konow, James. 2001. "Fair and Square: The Four Sides of Distributive Justice." *Journal of Economic Behavior and Organization* 46 (2): 137–64.

Laroque, Guy R. 2005. "Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof." *Economics Letters* 87 (1): 141–44.

Layard, Richard. 2005. *Happiness: Lessons from a New Science*. London: Penguin.

Lewbel, Arthur. 1997. "Consumer Demand Systems and Household Equivalence Scales." In *Handbook of Applied Econometrics; Volume 2; Microeconomics*, edited by M. Hashem Pesaran and Peter Schmidt, 167–201. Oxford: Blackwell.

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl. 2017. "Taxation and the Allocation of Talent." *Journal of Political Economy* 125 (5): 1635–82.

Lockwood, Benjamin B., and Matthew Weinzierl. 2012. "*De Gustibus non est Taxandum*: Heterogeneity in Preferences and Optimal Redistribution." National Bureau of Economic Research Working Paper 17784.

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

Maniquet, François, and Yves Sprumont. 2004. "Fair Production and Allocation of an Excludable Nonrival Good." *Econometrica* 72 (2): 627–40.

Maniquet, François, and Yves Sprumont. 2005. "Welfare Egalitarianism in Non-rival Environments." *Journal of Economic Theory* 120 (2): 155–74.

Maniquet, François, and Yves Sprumont. 2010. "Sharing the Cost of a Public Good: An Incentive-Constrained Axiomatic Approach." *Games and Economic Behavior* 68 (1): 275–302.

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

Mankiw, N. 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, N. Gregory, Matthew Weinzierl, and Danny Yagan. 2009. "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives* 23 (4): 147–74.

McQuillin Ben, and Robert Sugden. 2012. "Reconciling Normative and Behavioural Economics: The Problems to be Solved." *Social Choice and Welfare* 38 (4): 553–67.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Mirrlees, James A. 1974. "Notes on Welfare Economics, Information and Uncertainty." In *Essays in Equilibrium Behavior Under Uncertainty*, edited by Michael Balch, Daniel McFadden, and Shih-yen Wu, 243–58. Amsterdam: North-Holland.

Mirrlees, James A. 1976. "Optimal Tax Theory: A Synthesis." *Journal of Public Economics* 6 (4): 327–58.

Mirrlees, James A. 1982. "The Economic Uses of Utilitarianism." In *Utilitarianism and Beyond*, edited by Amartya K. Sen and Bernard Williams, 63–84. Cambridge and New York: Cambridge University Press.

Mirrlees, James A. 1986. "The Theory of Optimal Taxation." In *Handbook of Mathematical Economics; Volume 3*, edited by Kenneth J. Arrow and Michael D. Intriligator, 1197–249. Amsterdam: Elsevier, North-Holland.

Mongin, Philippe, and Marcus Pivato. 2016. "Social Evaluation under Risk and Uncertainty." In *Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, 711–45. Oxford and New York: Oxford University Press.

Moulin, Herve. 1987. "Egalitarian-Equivalent Cost Sharing of a Public Good." *Econometrica* 55 (40): 963–76.

Murphy, Liam, and Thomas Nagel. 2002. *The Myth of Ownership: Taxes and Justice*. Oxford and New York: Oxford University Press.

Negishi, Takashi. 1972. *General Equilibrium Theory and International Trade*. Amsterdam: Elsevier, North-Holland.

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

Pazner, Elisha, and David Schmeidler. 1978. "Egalitarian-Equivalent Allocations: A New Concept of Economic Equity." *Quarterly Journal of Economics* 92 (4): 671–87.

Piketty, Thomas, and Emmanuel Saez. 2013a. "A Theory of Optimal Inheritance Taxation." *Econometrica* 81 (5): 1851–86.

Piketty, Thomas, and Emmanuel Saez. 2013b. "Optimal Labor Income Taxation." In *Handbook of Public Economics, Volume 5*, edited by Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, 391–474. Amsterdam: Elsevier, North Holland.

Preston, Ian, and Ian Walker. 1999. "Welfare Measurement in Labor Supply Models with Nonlinear Budget Constraints." *Journal of Population Economics* 12 (3): 343–61.

Rawls, John. 1971. *A Theory of Justice*. Cambridge and London: Harvard University Press.

Rawls, John. 1982. "Social Unity and Primary Goods." In *Utilitarianism and Beyond*, edited by Amartya K. Sen and Bernard Williams, 159–86. Cambridge and New York: Cambridge University Press.

Rochet, Jean-Charles, and Lars A. Stole. 2003. "The Economics of Multidimensional Screening." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, edited by Matthias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky. Cambridge and New York: Cambridge University Press.

Robbins, Lionel. 1935. An Essay on the Nature and Significance of Economic Science, Second revised edition. London: Macmillan, 1937.

Roemer, John E. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy and Public Affairs* 22 (2): 146–66.

Roemer, John E. 1996. *Theories of Distributive Justice*. Cambridge and London: Harvard University Press.

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

Roemer, John E. 2008. "Harsanyi's Impartial Observer Is Not A Utilitarian." In *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, edited by Marc Fleurbaey, Maurice Salles, and John A. Weymark, 129–35. Cambridge and New York: Cambridge University Press.

Roemer, John E., et al. 2003. "To What Extent Do Fiscal Regimes Equalize Opportunities for Income Acquisition among Citizens." *Journal of Public Economics* 87 (3–4): 539–65.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68: 205–29.

Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses." *Quarterly Journal of Economics* 117 (3): 1039–73.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." *American Economic Review* 106 (1): 24–45.

Salanie, Bernard. 2011. *The Economics of Taxation, Second edition*. Cambridge and London: MIT Press.

Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Cambridge and London: Harvard University Press.

Samuelson, Paul A. 1974. "Complementarity—An Essay on the 40th Anniversary of the Hicks–Allen Revolution in Demand Theory." *Journal of Economic Literature* 12 (4): 1255–89.

Sandmo, Agnar. 1993. "Optimal Redistribution When Tastes Differ." *FinanzArchiv/Public Finnce Analysis* 50 (2): 149–63.

Sen, Amartya. 1985. *Commodities and Capabilities*. Amsterdam: Elsevier, North-Holland.

Shafir, Eldar. 2016. "Preference Inconsistency: A Psychological Perspective." In Oxford *Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, 844–70. Oxford and New York: Oxford University Press.

Sheffrin, Steven M. 2013. *Tax Fairness and Folk Justice*. Cambridge: Cambridge University Press.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Sprumont, Yves. 2013. "On Relative Egalitarianism." *Social Choice and Welfare* 40 (4): 1015–32.

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics* 17 (2): 213–40.

Stiglitz, Joseph E. 1987. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." *Handbook of Public Economics, Volume 2*, edited by Alan J. Auerbach and Martin Feldstein, 991–1042. Amsterdam: Elsevier, North-Holland.

Valletta, Giacomo. 2014. "Health, Fairness and Taxation." *Social Choice and Welfare* 43 (1): 101–40.

Van Praag, Bernard M.S., and Ada Ferrer-i-Carbonell. 2008. *Happiness Quantified: A Satisfaction Calculus Approach. Second Revised Edition*. Oxford and New York: Oxford University Press.

Varian, Hal R. 1974. "Efficiency, Equity and Envy." *Journal of Economic Theory* 9 (1): 63–91.

Varian, Hal R. 1976. "Two Problems in the Theory of Fairness." *Journal of Public Economics* 5 (3–4): 249–60.

Varian, Hal R. 1980. "Redistributive Taxation as Social Insurance." *Journal of Public Economics* 14 (1): 49–68.

Vickrey, William. 1945. "Measuring Marginal Utility by Reactions to Risk." *Econometrica* 13 (4): 319–33.

Weinzierl, Matthew. 2011. "The Surprising Power of Age-Dependent Taxes." *Review of Economic Studies* 78 (4): 1490–518.

Weinzierl, Matthew. 2012. "Why Do We Redistribute so Much but Tag so Little? Normative Diversity, Equal Sacrifice and Optimal Taxation." National Bureau of Economic Research Working Paper 18045.

Weinzierl, Matthew. 2014a. "Revisiting the Classical View of Benefit-Based Taxation." Harvard Business School Working Paper 14–101.

Weinzierl, Matthew. 2014b. "The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice." *Journal of Public Economics* 118: 128–42.

Wilson, Robert B. 1993. *Nonlinear Pricing*. Oxford and New York: Oxford University Press.

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

Young, H. Peyton. 1987. "Progressive Taxation and the Equal-Sacrifice Principle." *Journal of Public Economics* 32 (2): 203–14.

Zelenak, Lawrence. 2006. "Taxing Endowment." *Duke Law Journal* 55: 1145–81.