Optimal taxation theory and principles of fairness*

Marc Fleurbaey† François Maniquet‡

Feb. 2015

Abstract

The achievements and limitations of the classical theory of optimal labor-income taxation based on social welfare functions are now well known, although utilitarianism still dominates public economics. We review the recent interest that 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 Classification: H21, D63.

Keywords: optimal taxation, fair social orderings.

*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 comments and suggestions by P. Pestieau, six referees, and the Editor. 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.

†Princeton University. Email: mfleurba@princeton.edu.

‡CORE, Université catholique de Louvain. Email: francois.maniquet@uclouvain.be.

1 Introduction

The theory of optimal income taxation has reached maturity and excellent reviews of the field are available (Mankiw et al. 2009, Salanié 2011, Piketty and Saez 2012). At the same time, insatisfaction appears to be growing about the difficulty of the theory to solve problems that have recently been raised.

A first source of insatisfaction comes from the fact that most of the corpus of optimal tax theory assumes that individuals have identical preferences. For instance, Boadway lists the "heterogeneity of individual utility functions" (2012, p. 30) as one of the big challenges for optimal tax theory (along 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" (p. 30-31).

The issue of making interpersonal comparisons of welfare is indeed much more than an issue of analytical simplicity. When individuals have the same utility function, the only 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.

There comes the second source of insatisfaction. 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 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

[1] A (constrained or unconstrained) Pareto-efficient allocation is an extremum 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). [footnote added]

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." (2012, p. 2) 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 idea that tagging[2] on the basis of statistical discrimination may violate a form of horizontal equity; the libertarian view that the distribution of earnings may deserve some respect; the principle that income inequalities due to differences in preferences are not as problematic as inequalities due to differences in qualifications or social background.

The first objective of this paper is to review and discuss the difficulties of classical optimal taxation theory, especially in its attempts to take preference heterogeneity into account. Although its limitations are now quite well known, it is worth carefully listing them and have them in the background for the examination of other frameworks. There is a temptation in the literature to throw the baby with the bath water, and our main message is that the classical framework should be extended rather than abandoned.

The second objective of the paper is to review some recent contributions that have invoked fairness principles to derive conclusions about income taxation. We discuss the extent to which such contributions solve the difficulties faced by classical optimal taxation theory. Along the way, we attempt to 1) clarify some possible misunderstanding about the compatibility between fairness principles and the Pareto principle; 2) provide intuitive explanations for notions that the literature often derives from axiomatic analysis; and 3) explain why one finds the maximin aggregator in several fairness approaches.

Our discussion will lead us to argue that fairness concepts can help solve difficulties of the classical approach to optimal income taxation theory not by overruling the classical social welfare functions, but by providing useful selections of such functions and in particular of suitable individual utility indexes. In particular, there is a way to construct utility functions embedding ethical principles for which the various sources of income inequalities do not equally call for redistribution.

[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.

In this paper we also review contributions that try to refine classical optimal tax theory by only amending the aggregator, introducing weights into the utilitarian social objective. This approach offers an alternative way to introduce fairness principles in optimal taxation. Compared to constructing suitable utility functions, however, introducing weights turns out to be less tractable and delivers recommendations that are only consistent with a narrower set of fairness principles.

Our paper is complementary to a recent paper by Saez and Stantcheva (2013). They propose to go beyond the classical social welfare function framework and to derive optimal taxes from the application of marginal social welfare weights directly to earning levels. The weights at each earning level depend on the characteristics of the agents earning that much, and can be inspired by fairness principles. In this fashion, Saez and Stantcheva are able to retrieve and extend some of the fair tax results. Compared to their approach, our contribution is to show that many relevant fairness considerations can actually be accommodated in the classical social welfare framework. Moreover, as they note, determining the weights to be applied to earnings is not always immediate from the reading of fairness principles and may require a detour which involves writing down the social welfare function. In a nutshell, we propose to broaden the considerations that shape the social welfare function rather than abandon the social welfare framework itself. But with them and the authors quoted in the beginning of this introduction we share the general goal of incorporating a broader set of ethical principles in optimal tax theory, and their analysis in terms of weights on earnings is definitely useful, as we will illustrate in this paper.

In the following sections, we begin with a brief description of the main achievements of the classical approach (section 2). We then discuss the difficulties associated with utilities as a proxy for well-being (section 3) and with utilitarianism as an aggregator of well-being levels (section 4). We review various fairness approaches to optimal taxation in section 5: Mankiw's "just deserts" approach, Roemer's equality of opportunity, the fair social ordering approach, and the luck-desert distinction discussed in Saez and Stantcheva (2013). In that section we also briefly discuss Kaplow and Shavell's (2001) challenge to fairness principles. In the following section, we discuss various attempts to incorporate fairness principles in a weighted utilitarian social welfare function (section 6). Next, we analyse the derivation of fair optimal tax and the usefulness of Saez and Stantcheva's (2013) approach in terms of marginal social welfare weights (section 7). Finally, we provide a simple methodology for linking the construction of utility functions with four connected but distinct ethical choices: subjective utility versus fairness, redistribution versus laissez-faire, compensation versus responsibility, and the relative treatment of individuals with different preferences (section 8). 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. We conclude in section 9.

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. 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 own analysis.

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, non-increasing 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 wage rate, 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 pre-tax income (earnings).

An allocation is a collection of bundles $z = (z_1, ..., z_n)$. A tax function $T : \mathbb{R}_+ \to \mathbb{R}$ 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 $i, j \in N$,

$$U_i(\ell_i, c_i) \geq U_i(y_j / w_i, c_j) \quad \text{or} \quad y_j > w_i.$$

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 constraint. Since Mirrlees (1971), the evaluation of T is derived from an 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 $((U_1, ..., U_n), (w_1, ..., w_n))$, defines a specific ordering (i.e., a complete transitive relation) on the set of allocations X^n. We retain this approach in all the paper.

The classical theory of labor income taxation has been initially developed under two main assumptions. First, agents in the economy have different productivity levels, but they all have the same preferences over labor-consumption bundles, represented by a single utility function: for all $i \in N$: $U_i = U_0$. Second, the social planner is utilitarian, which means that the social ordering function is defined as maximizing the sum of utility levels:

$$\sum_i U_0 (z_i) . \tag{1}$$

A more general social welfare function has also been considered, but under separability assumptions 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, the design of second-best tax schemes, and the 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).

In the last fifteen years, the theory has been enlarged to consider the more realistic case in which agents also differ in their preferences. Introducing additional dimensions of heterogeneity in the picture 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 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 cases (see, e. g. Boadway, Marchand, Pestieau and Racionero 2002, Jacquet and Van de Gaer 2011, Choné and Laroque 2012). A general solution has also been proposed by Saez (2001, 2002), recently refined and extended in Jacquet and Lehmann (2014). Saez's approach consists in 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.

The income weight approach offers a valuable solution to the technical difficulties of optimal tax theory in the presence of heterogenous preferences.[3] Nonetheless, the question of how to make interpersonal utility comparisons, and, more specifically, how to compare high-productivity-low-willingness-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 are utilities?

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.

[3]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).

There are two main views on utilities. According to the first view, 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 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. This has been popularized in the last two decades by the economics of happiness. It builds on answers to survey question 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 to it.[4]

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 (1981), 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 et al. (2009) 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.[5]

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

[5] A much more extensive discussion of the subjective well-being approach can be found

Philosophers have suggested to replace utilities with other arguments. Dworkin, 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" (p. 94-95). The same authors also report the 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-word opinions" (p. 137-138). As we will see in section 5, these questions are at the heart of the fairness approaches to optimal taxation. The fairness principles discussed later in this paper receive a substantial support in the empirical literature surveyed in Gaertner and Schokkaert (2012).

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 against this family of theories. The first criticism opposes the assumption that risk aversion is a measure of intensity of preferences (Roemer 1996). 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

in Fleurbaey and Blanchet (2013, chapter 5).

social criterion, and this is especially relevant when one deals with heterogenous preferences (which was not Mirrlees' frame). 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 (Luce and Raiffa 1957, Dhillon and Mertens 1993, Sprumont 2010, Adler 2014). 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 2015). Our review of the criticisms is meant to prove one point: Adopting such measures cannot be done without relying on strong ethical assumptions. They are not neutral and ready-to-use measures.[6]

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 directly compa-

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

rable. 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. The classical approach to optimal taxation has not explored how to construct utilities in this perspective. Fairness approaches are meant to fill this gap, as we will illustrate in sections 5 and 8 below.

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. Independently of the choice of utility functions that are used in this summation, we can list four shortcomings of the utilitarian aggregator in the context of optimal income taxation.

The first shortcoming 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. This is in sharp contrast with what ethical intuition would recommend in that particular case. One may, indeed, argue that differences in productive abilities do not justify differences in outcomes. This calls for equalizing satisfaction levels among agents with the same preferences. Another view holds that agents own their ability at least partially, so that the high ability agents should reach a higher satisfaction level. The utilitarian objective is unable to produce either result.

The second shortcoming 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.

The third shortcoming is emphasized by Piketty and Saez (2012). Maximizing a sum of utilities, even weighted,[7] cannot produce the desirable properties that 1) utility should be equalized when all agents have the same

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

preferences, an objective which we will refer to later as the *compensation* objective,[8] and 2) laissez-faire should prevail when all agents have the same productivity level, an objective which we will refer to later as the *laissez-faire* objective.[9] 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 ethical goals of equalizing utilities among agents having the same preferences and not redistributing among agents having the same productivity are, on the other hand, at the heart of several fairness approaches.[10] In the next section, we illustrate several ways of combining these goals and we show how they can be used to define social objectives that are ethically grounded.

The fourth shortcoming is related to tagging. 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 dimensions that are ethically relevant but are different with respect to the dimension along which people are tagged, they may be treated differently and end up enjoying different satisfaction levels.

If the social criterion is utilitarian, the sum of the utilities will necessarily increase, but at the cost of a gap in the utilities of agents who should be considered equal. The shortcoming of utilitarianism in this respect is that the gap between the utilities of these two agents can take place at the expense of the worse off agent. That is, the agent with the unfavorable situation may end up at a lower utility level than without 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

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

[9]Jacquet and Van de Gaer (2011) prove a similar statement, restricting their attention to two planner objectives, the non-weighted sum of utilities and the non-weighted sum of a concave transformation of the utilities.

[10]In the literature, the former goal is often referred to as the compensation principle, whereas the latter goal is referred to as the responsibility principle, or liberal reward principle (see, e. g. Fleurbaey, 2008, and Fleurbaey and Maniquet, 2011, for surveys of the literature on these two principles).

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. It is unlikely, though, that the tall unskilled ones will benefit as well.

Let us note that this cannot happen if the criterion is strongly egalitarian, such as a maximin criterion. As soon as the satisfaction of the social criterion increases, by assumption, it cannot be obtained at the expense of the worse off. Of course, tagging still leads under the maximin to violations of the principle of equal treatment of equals, but the violations cannot happen among the worse off agents. Concretely, the unskilled, whether small or tall, need to end up enjoying the same increased satisfaction level.

More generally, the maximin criterion, applied to utilities, independently of the way utilities are built, escapes three of the four shortcomings that we just mentioned. As we will see in the next section, the maximin criterion plays an important role in the fairness approach to optimal taxation, although it is applied to well-being measures that are different from the ones that are typically used in classical optimal taxation theory.

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 egalitarian) requirements on other objects than utilities. Before we begin this review, though, we need to clarify the relationship between our notion of fairness and the Pareto principle.

Kaplow and Shavell (2001) have argued that any continuous and non-welfarist method of policy assessment violates the Pareto principle. In their work, welfarism is defined by the axiom of Pareto indifference: as soon as all agents are indifferent between two allocations, society should also be indifferent between these two allocations. These authors were targeting normative theories proposing to discriminate among Pareto indifferent allocations on the basis of fairness principles.

All the fairness approaches that we review in this section satisfy the Pareto principle. In our terminology, though, they are not welfarist, because we stick to the classical definition of welfarism in welfare economics. Welfarism, indeed, is the theory that utilities should be aggregated in the same way (the utilitarian way, for instance, i.e., summing up utility levels) independently of the profile of utility functions in the population. Classical optimal income taxation has always been consistent with this definition of welfarism, and the difficulties we have listed in the previous two sections are closely related to utilitarianism being a welfarist theory of justice.

5.1 Libertarianism

In a recent contribution, Mankiw (2010) revives the libertarian view on labour income taxation. He advocates a principle of Just Deserts. The general principle is that “a person who contributes more to society deserves a higher income that reflects those higher contributions.” In the absence of market imperfections, “each person’s income reflects the value of what he contributed to society’s production of goods and services.” Let us note that the absence of market imperfections is the typical assumption of optimal tax theory, as the wage rates are assumed to be equal to the real productivity of the agents. Mankiw also argues that market imperfections, such as pollution, or the provision of public goods, such as fighting poverty, should be funded according to the agents’ incomes, because richer agents benefit more from them. As a result, a progressive tax scheme can still emerge from the libertarian view.

Maybe the most radical departure from the utilitarian approach to taxation, in Mankiw’s approach, is that the issue of how to make interpersonal comparisons has disappeared from the picture. More precisely, constructing comparable utilities has been replaced with the application of an extended version of the laissez-faire objective: earnings are fair if they reflect the natural differences among people, and these differences come from differences in talents, which are rewarded at their marginal productivity, and differences in preferences, which are rewarded proportionally to labor times. Inequalities due to differences in earning capacities are no longer considered unjust.

This is a rather extreme postulate, in the spectrum of fairness theories. It does not receive much support from popular views on justice. Using questionnaire surveys to elicit ethical preferences, Konow (2001) finds support for the view that “a worker who is twice as productive as another should be paid twice as much if the higher productivity is due to greater work effort but not if it is due to innate aptitude” (p. 138).

However, the main point we want to make here is that with a suitable choice of utility functions one can construct a social welfare function that seeks to achieve the laissez-faire allocation. Therefore, even libertarianism can be accommodated, at least to some extent, in the social welfare framework. 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 (note that in the following formula, as well as later on in the paper, we use t to denote lump sum transfers) the value of the expenditure function for a reference wage rate w and the utility level $U_i(z_i)$:

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

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

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

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 the best feasible allocation. Moreover, this ordering is intuitive because it considers that the worse-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

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

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. This again illustrates the key idea of this paper, namely, that the choice of utility indexes is the important factor

One could object that the same feat can be achieved by a weighted sum of utilities, for a suitable choice of weights. While it is true (under mild assumptions) that the laissez-faire allocation maximizes a weighted sum of utilities, 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 into any economy with an arbitrary profile, and any inequality aversion social welfare function, it makes the laissez-faire allocation the best. The limitations of weighted utilitarianism will be discussed further in section 6.

5.2 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). The central idea of the theory is that the sources of inequalities in individuals' achievements should be divided in 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. The measurement of effort therefore requires partitioning the population by circumstances (i.e., by types), and measuring effort within each type by the relative ranks in the distribution of outcome.

Roemer et al. (2003) apply the criterion to optimal 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.

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 first objective 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.

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 gives an easy but superficial solution to the question of the 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 use his approach to utilities and has restricted the applications of his criterion to cases in which the outcomes naturally come in cardinal and comparable 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 Mankiw's 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 would have the same circumstances, including the same skills. The just desert approach would recommend to let agents be rewarded according to their skill, and all income differences would come from different choices, for which no correction should be implemented. Laissez-faire is considered fair. Roemer's equality of opportunity approach would recommend indifference about inequalies 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 compatible with income redistribution even in economies in which all agents have the same earning capacities.

5.3 The fair social ordering approach

The third fairness approach we survey offers a combination of the first two approaches. This approach, indeed, pursues Roemer's compensation objective, under the assumption that the effort characteristics are the preferences of the agents. The 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 —see below). 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.[11]

Let us note that by combining Roemer's 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.[12]

It may be useful at this point to clarify the definition of the (pairwise) compensation objective and avoid a possible mistake. The compensation principle requires that two agents having the same preferences should enjoy the same satisfaction level. It should be clear that this principle remains a purely ordinal one. Enjoying the same satisfaction level, indeed, means that these two agents should consume bundles they deem equivalent. Equivalently, the requirement can be stated by reference to the celebrated (ordinal) fairness concept of no-envy, introduced in the formal theory of fair allocation by Kolm (1972) and Varian (1974): such agents should not envy each other.[13]

In this section, we illustrate the combination of the compensation and the laissez-faire objectives by introducing an important class of social orderings. The precise underlying ethical values, and other possible orderings embodying different value judgments, are 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

[11] 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).

[12] However, they also obtain results which 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?

[13] For a detailed study of the relationship between no-envy on the one hand and the compensation and the liberal reward principles in the other hand, see Fleurbaey (2008) and Fleurbaey and Maniquet (1996).

at a common reference wage $\tilde{w}$. This social ordering is then represented by the function[14]

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

For the moment, let us only assume that the reference wage lies between the lowest and the largest wage rates 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 wage rates 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[15] social ordering functions.

The first point we want to make here is that, in a first best world,[16] 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.

A frequent criticism against this form of egalitarianism is 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 address this criticism, 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

[14] 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.

[15] 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 no tax and a wage rate equal to $\tilde{w}$.

[16] The first best world is also the second best world when preferences are special and make agents work the same labor quantity for all tax functions.

directly connects to the discussion of "expensive tastes" and adaptation in section 3. It is clear 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. If the criticism that our compensation objective does not take these inequalities sufficiently into account is based on this fact, it means that the model is incomplete and such internal parameters have to be made explicit, together with individual preferences over them. Note that agents would then have three sources of heterogeneity in such an extended model.[17] In conclusion, either way, the criticism can be suitably addressed. We continue the discussion, though, 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, necessarily $\tilde{w}$ must equal the common wage rate, and equality of $m_i(\tilde{w}, z_i)$ 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 $m_i(\tilde{w}, z_i)$ 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.[18] 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

[17] Such a triple-heterogeneity model is studied by Fleurbaey and Valletta (2014). The problem of compensating for unequal internal characteristics is developed at length in a related literature (Fleurbaey 2008, Fleurbaey and Maniquet 2011b, 2012).

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

curve in X lies everywhere above another agent's. In this case, both agents agree that the bundle of goods consumed by the former agent is preferable to the bundle consumed by the latter. Moreover, they both also agree that any bundle that the former agent finds indifferent to the bundle he consumed is preferable to any bundle that the latter agent finds indifferent to the bundle he consumes. In that case, the money-metric utility of the latter is necessarily lower than the former'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.[19] It applies to individuals having the same wage rate, but only when their common wage rate 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 $m_i(\tilde{w}, z_i) = m_j(\tilde{w}, z_j)$ 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 rate. This partial laissez-faire outcome is directly related to the way utilities are constructed. For individuals with a common wage rate $w_i = w_j$ that does not differ too much from $\tilde{w}$, this laissez-faire property will only hold approximately, i.e., the equality $m_i(w_i, z_i) = m_j(w_j, z_j)$ 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.

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 wage rates) 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).

[19]The fact that one property (the compensation principle) is satisfied to a larger extent than the other (the laissez-faire) is due to a logical tension between the two. This is further discussed in Section 8.

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, Chap. 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) can be satisfied only with an absolute priority for the worst-off when the evaluation satisfies informational simplicity requirements.

The reasoning behind this absolute priority to the worst-off goes as follows. The Pareto criterion forces us to aggregate satisfaction levels or indifference curves, which means that not too much attention can be paid to the bundles of goods consumed by the agents. On the other hand, the fairness objectives that have been studied so far are defined as Pigou-Dalton transfer requirements: a transfer of resources from an advantaged agent to a relatively disadvantaged one is a social improvement. Consequently, fairness forces us to look at what agents consume. Combining the Pareto criterion with this transfer requirement turns out to be very demanding. This is typically the case when the shape of the preferences of the agents is such that the value for them of a given transfer depends on the bundle they originally consume when they receive that transfer. A transfer of a given amount of resources at a certain level of work may be, in terms of preferences satisfaction, equivalent at another level of work to the loss of an arbitrarily large amount of resources by the richer agent combined with the gain of an arbitrarily small amount of resources by the relatively poor one. It is intuitive that as a consequence, transfers are a social improvement only if we use a maximin aggregator.[20]

The maximin conclusion can be avoided by applying the Pigou-Dalton transfer requirements only to profiles of preferences that are sufficiently regular (e.g., homothetic or quasi-linear) in order to avoid the equivalence between balanced transfers and unbalanced transfers depicted in the previous paragraph. One then obtains social orderings with a less than absolute priority for the worst-off.

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, agents are characterized by their preferences and their skill. That forces us to put the cut between circum-

[20] However, a full argument requires an additional restriction on the information about indifference curves that can be used to evaluate a transfer, or a separability property for the evaluation of subpopulations. See Fleurbaey and Maniquet 2011a, Chap. 3 for details.

stances 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 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 (Choi et al. 2003).

5.4 Luck and desert

Saez and Stantcheva (2013), 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 is assumed to have a zero mean. They show in particular that if only the individuals with net income below their earnings are given a positive weight in the social objective, income taxation may have to be greater when the random shock has greater variance relative to post-tax earnings, creating a reinforcing mechanism that can generate multiple equilibria. Low-tax equilibria incur a lower random shock relative to post-tax earnings, justifiying the low tax, and conversely 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 fair social ordering approach.

In Roemer's approach, the random shock can be added to the circumstances, and if the random shock is uncorrelated with the pre-shock outcome 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 bad otherwise circumstances.

Roemer's criterion would then advocate maximizing their average outcome.

In the fair social ordering 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 the same 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 rate equal to $\tilde{w}$, the ideal state is to give them lump-sum transfers cancelling their inequalities in luck and let them work freely (with no further tax).

The advantage of this approach is that it provides a sensible ordering of individuals, that 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 over income is endogenous to the tax. In particular, if the tax is 100%, as they note, the post-tax earnings are null and all income is undeserved (and has zero elasticity), implying that the optimal tax is indeed 100%. In contrast, a social welfare function with $m_i(w_i, z_i)$ is unlikely to generate multiple equilibria and to be satisfied with a 100% tax. The reason is that it does not treat post-tax 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 Does weighting utilities yield fair outcomes?

The key building block of the reference-wage egalitarian-equivalent social ordering functions is the money-metric utility $m_i(\tilde{w}, z_i)$ which makes interpersonal comparisons of resources appropriate in the double way that was needed to achieve the desired combination of the compensation and laissez-faire objectives, that is, both among agents with identical preferences and agents with the reference wage rate. 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 rate 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".[21] The money-metric utility respects individual preferences while using an objective 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 rate.

It is remarkable how many possibilities are offered for social evaluation by an appropriate selection of well-being indexes, even under the restriction that the indexes must be utility functions that represent individual preferences, and even under the restriction that the maximin criterion will be ultimately applied to the distribution of such indexes. A substantial portion of this diversity of possibilities, and the underlying ethical issues, is illustrated in section 8.

[21] 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)

Before exploring what ethical principles may guide the selection of suitable indexes, it is worth examining a more basic question. If the population comes with a given profile of utility functions $(U_1, ..., U_n)$, is it necessary to replace such functions by indexes like $m_i(\tilde{w}, z_i)$ before applying a social welfare function (such as the maximin)? Couldn't one simply weight these utility functions in a utilitarian sum?

Facing the problem of heterogeneous utilities, the literature[22] has indeed considered weighting them in a utilitarian social welfare function $\sum_{i\in N} \alpha_i U_i(z_i)$. These coefficients α_i's are the so-called Pareto weights referred to in the introduction (quoting Piketty and Saez 2012). 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, 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 rate 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 rate will always have a higher well-being index $m_i(\tilde{w}, z_i)$ than an agent with the same preferences and the lowest wage rate. 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 rate, 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 α's. 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

[22] See in particular Boadway et al. (2002), Kaplow (2008), Choné and Laroque (2012).

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. As a result, new α's 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 α's could only be ascertained after the optimal allocation is identified.

We can apply these ideas to the recent interesting work of Lockwood and Weinzerl (2014). Following the simplifying method to deal with heterogenous preferences introduced by Brett and Weymark (2003), they assume that individual 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

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

i.e., the average value, among the agents of (actual) parameter n, of $g(.)$ in the laissez-faire allocation of a hypothetical economy in which all agents have the average wage rate $\bar{w}$ of the actual economy. When the actual economy already has equal wage rates 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

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

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.[23]

Whatever the precise formula for the weights, the evaluation of allocations in the *actual* economy is not geared toward the laissez-faire in any clear way. In particular, the weighted objective does 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 rate $\bar{w}$ but different θ_j, θ_k. The sum

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

has no reason to 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.

The problem would persist if one considered a weighted form of maximin in which the same exogenous utility functions are weighted before the maximin criterion is applied to them. Therefore, it does not appear very promising to rely solely on weighting utilities in order to enrich the classical social welfare function. What is needed is a more sophisticated choice of an appropriate representation of individual preferences.

The replacement of arbitrary utility functions $U_i(z_i)$ by suitable well-being indexes is therefore not a superfluous exercise that could be mimicked by a weighting system. Weighted utilitarianism is not the all-purpose tool that it is often believed to be. Relying on it in order to incorporate the

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

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 (2013). 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 function T will induce a change in social welfare equal, by the envelope theorem,[24] to

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

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. A difficulty with this approach is that the weights are endogenous and depend on the particular allocation under consideration. In this section, we show how the theory of fair social orderings can help compute

[24] The additional term

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

vanishes when either the first-order condition

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

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

these weights.[25]

A simplification comes from the fact that, given that fair social orderings are of the maximin type, the maximal weight will 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 two such cases. In both cases, we assume that the planner is interested in "maximinning" a particular well-being index in the reference-wage egalitarian-equivalent family presented above, $m_i(w_{\min}, z_i)$, where $w_{\min} = \min_{i \in N} w_i$.

The first case is a reform context. Let us assume that a tax scheme prevails but is not optimal. A reform has to be designed, but the tax scheme can only be changed marginally. How should we change it?

The reasoning is illustrated in Fig. 1. It represents the pre-tax/after-tax income space. The 45° line represents the relation between pre-tax and after-tax incomes in the absence of taxation. The tax scheme that we try to evaluate, T, is represented by the corresponding function that describes how after-tax incomes, c, depend on pre-tax incomes, y: $c = y - T(y)$.

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 earn an income in the $[0, w_{\min}]$ interval (remember that w_i also stands for agent i's income, would he work full time).

Let $y^* \in [0, w_{\min}]$ be the pre-tax income for which the tax $T(y)$ is maximal over the $[0, w_{\min}]$ interval. Graphically, y^* is the abscissa of the point of

[25] Saez and Stantcheva (2013) provide many examples covering tagging, desert, fairness. They extend the approach to cases in which the weights are not related to an underlying social welfare function, which enlarges the set of possible ethical approaches covered by their analysis. We restrict attention here to weights which derive from a social ordering (i.e., transitivity and the Pareto principle are respected).

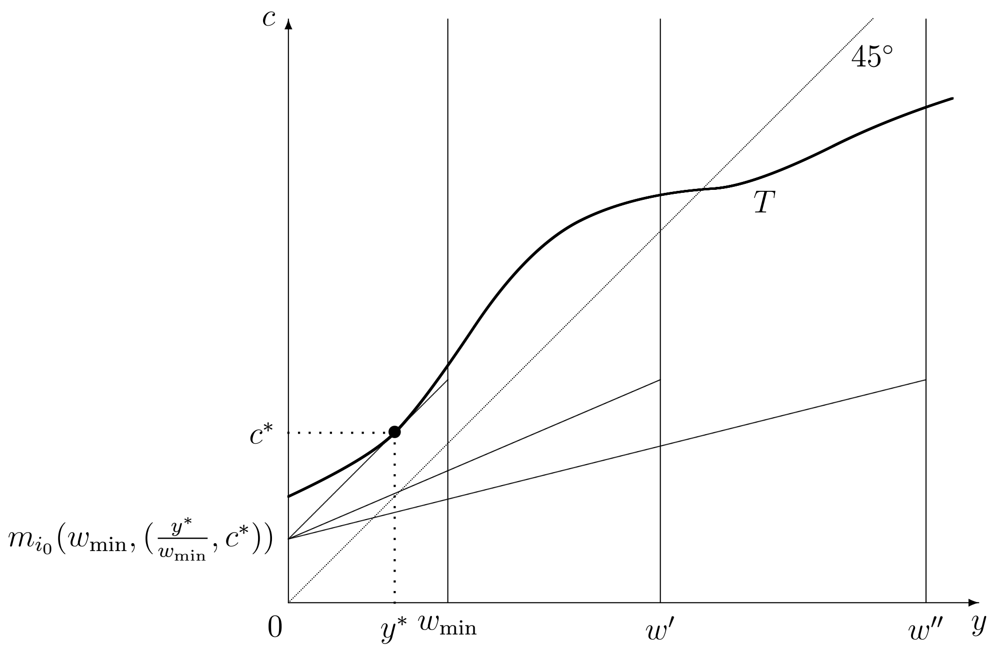

Figure 1: Deriving the income weights: the case of a reform.

tangency between the curve representing T and a 45° line.[26] Let us assume that some agents with minimal productivity, one of them having index i_0, earn an income of y^*.[27] Given the tax scheme, this allows them to reach an after-tax income of c^*.

Individual i_0's choice reveals that he prefers bundle $(\frac{y^*}{w_{\min}}, c^*)$ to all other possible bundles affordable given the tax scheme. The segment of a 45° line that is tangent to the function describing T is also drawn in the figure. We claim it represents the implicit budget of agent i_0, i.e., the budget $c = t + w_{\min}\ell = t + y$ that underlies the computation of $m_{i_0}(w_{\min}, z_{i_0})$. Indeed,

[26]In this example, y^* is interior to the interval. It need not be the case. It is even quite likely to have $y^* = w_{\min}$.

[27]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}.

the slope of that budget, from the point of view of agent i_0, is precisely $w_{\min}$, because it is her actual productivity, so that it corresponds to her marginal pay in the absence of taxation. Consequently, the intercept of that segment measures the value of the well-being index of agent i_0 at that bundle, $m_{i_0}(w_{\min},(\frac{y^*}{w_{\min}},c^*))$.

It is impossible, on the other hand, to draw the implicit budgets of the agents having their productivity equal to w_{min} but choosing other bundles: their choice does not reveal sufficient information for us to identify their indifference curve. What we can say about them, however, is that their choice reveals that they prefer their bundle not only to $(\frac{y^*}{w_{\min}},c^*)$, but also to any bundle in the budget line determined by $w_{\min}$ and containing the bundle (y^*,c^*), because this budget line is never above the budget curve delineated by the tax. As a consequence, their implicit budget is not below agent i_0's. This just proves that the agents choosing to earn y^* among the lowest productivity agents have the lowest well-being index.

Second, we need to compare the well-being index among agents of different productivity subgroups. Given that we only need to identify who has the lowest well-being index in the population, we can focus on agents having the same or a lower well-being index than agent i_0.

Two more lines are drawn in the figure, starting from the intercept $m_{i_0}(w_{\min},(\frac{y^*}{w_{\min}},c^*))$. They represent the implicit budget of agents having their productivity equal to w' and w'' respectively. Their slope is lower than that of the implicit budget of agent i_0. Indeed, productivity $w_{\min}$ is not their actual productivity. The implicit budgets should tell us how much they would earn, should their productivity be equal to $w_{\min}$, which is a fraction $w_{\min}/w'$ (resp., $w_{\min}/w''$) of their actual productivity. These fractions are the precise slopes of the two lines.

The figure shows that these two implicit budgets are strictly below the function describing the tax scheme. It is a necessary consequence of the fact that this function is increasing. This means that all agents having a greater productivity than $w_{\min}$ prefer their actual bundle to being allowed to maximize their preferences over implicit budgets of intercept $m_{i_0}(w_{\min},(\frac{y^*}{w_{\min}},c^*))$. As a result, all agents having a productivity equal to w' or w'', and, more generally, greater than $w_{\min}$, have an implicit budget above agent i_0's. The latter is therefore the agent with the lowest well-being index. The egalitarian planner should therefore assign a positive weight to income y^* and a zero weight to all other incomes that are strictly above the 45° line segment that is tangent to the budget at y^*.

The second case in which it is possible to derive income weights is when the planner tries to maximin the well-being index $m_i(w_{\min}, z_i)$ and the allocation is second-best optimal for these social preferences. More precisely, we will show that the marginal rate of taxation should be zero on incomes below $w_{\min}$, and all income weights should be zero above $w_{\min}$. This gives us a precise formula for the optimal tax scheme for those social preferences. This case is also discussed in Saez and Stantcheva (2013), and we will provide an intuitive explanation for the zero marginal tax result.

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

By the same reasoning as for the first case above, we can identify the level of the lowest well-being index in the population. It is the intercept of the lowest implicit budget that is tangent to the budget function representing T from below in the $[0, w_{\min}]$ range of incomes. It is denoted b in the figure. Formally,

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

Observe that all agents earning an income of y' or less have a subsidy greater than 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'. Their influence on the budget 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 tax payer 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 which 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 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.

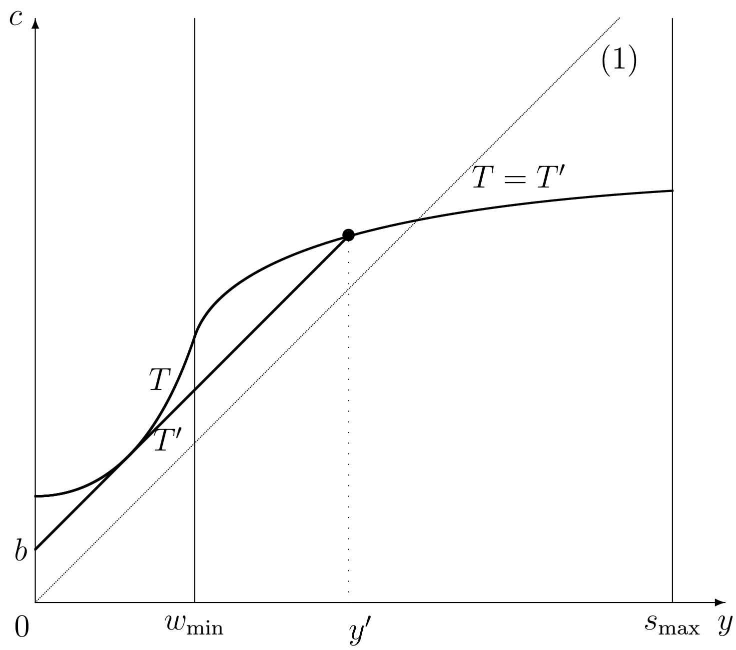

Figure 2: Optimal allocations for $\mathbf{R}^{w_{\min}\text{lex}}$ and $\mathbf{R}^{\tilde{\ell}EW}$ have a zero marginal rate on low incomes.

How should incomes be taxed above $w_{\min}$? As we proved in the first part of this section, the well-being index of agents earning more than $w_{\min}$, and, therefore, having a larger productivity than $w_{\min}$, cannot be lower than the well-being index of the agents having the lowest productivity. The only objective of taxing those incomes should then be to maximize the tax return so as to have as large a subsidy on low incomes as possible (under the constraint, of course, that after-tax income remains an increasing function of pre-tax 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[28]

$$\begin{aligned} \forall y \leq w_{\min}, &\quad T'(y) = 0 \\ \forall y \geq w_{\min}, &\quad \frac{T'(y)}{1-T'(y)} = \frac{1-F(y)}{\epsilon(y) y f(y)}. \end{aligned}$$

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 in 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 weights is not as 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 a short-cut, unfortunately, for the determination of the optimal tax. Note also that 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. This belongs to the set of open questions for future research.

8 Four ethical choices

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

[28] 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 functon. See Saez (2001) for the derivation of this formula, and Jacquet and Lehmann (2014) for a similar formula.

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 a review of indexes in Preston and Walker (1994), where the underlying normative principles were not made explicit, and an analysis in Decoster and Haan (2014), where three main examples of indexes are discussed together with their ethical meaning.

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 them? 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.

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 the "expensive tastes" and the adaptation problem than subjective declarations of utility or satisfaction. Moreover, subjective declarations may contradict the individuals' own interpersonal comparisons, which are especially compelling when indifference curves do not cross. An individual may be on higher indifference curves but declare a lower satisfaction, simply because individual self-assessments rely on heterogenous 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 non-comparable preferences. In the rest of this section we focus on indexes that are based on ordinal non-comparable preferences in order to show the wide array of possibilities offered by this informational basis.

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 between 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 $m_i(\lambda\tilde{w} + (1-\lambda) w_i, z_i)$, in order to let individuals enjoy their skills to some extent $(1-\lambda)$ and contain 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).

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

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

where $\tilde{w}_i = \lambda w_i + (1-\lambda)\tilde{w}$. 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 because the implicit budgets for greater skills are then dominated by the implicit budgets for lower skills (except at $\ell = 1$, where they meet).

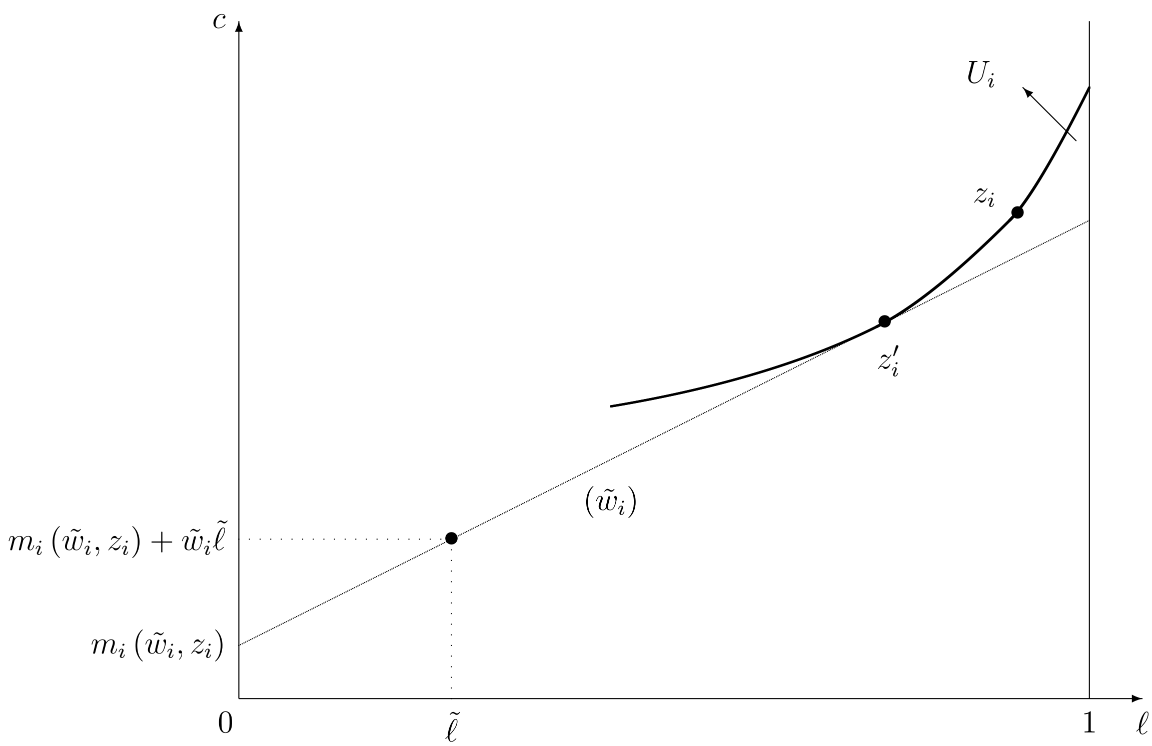

Figure 3: The construction of well-being indices:

A further generalization (explored in Fleurbaey and Maniquet 1996) 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.[29] In order 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 $\tilde{w}_i = \lambda w_i + (1-\lambda)\tilde{w}$: 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.

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 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 rate 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 rate, 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 rate. 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 wage rates 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 $m_i(\tilde{w}, z_i)$ and $m_i(w_i, z_i) + w_i\tilde{\ell}$ is a choice between compensating unequal skills and laissez-faire among individuals with identical skills: pairwise compensation versus pairwise laissez-faire.

[29] 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.

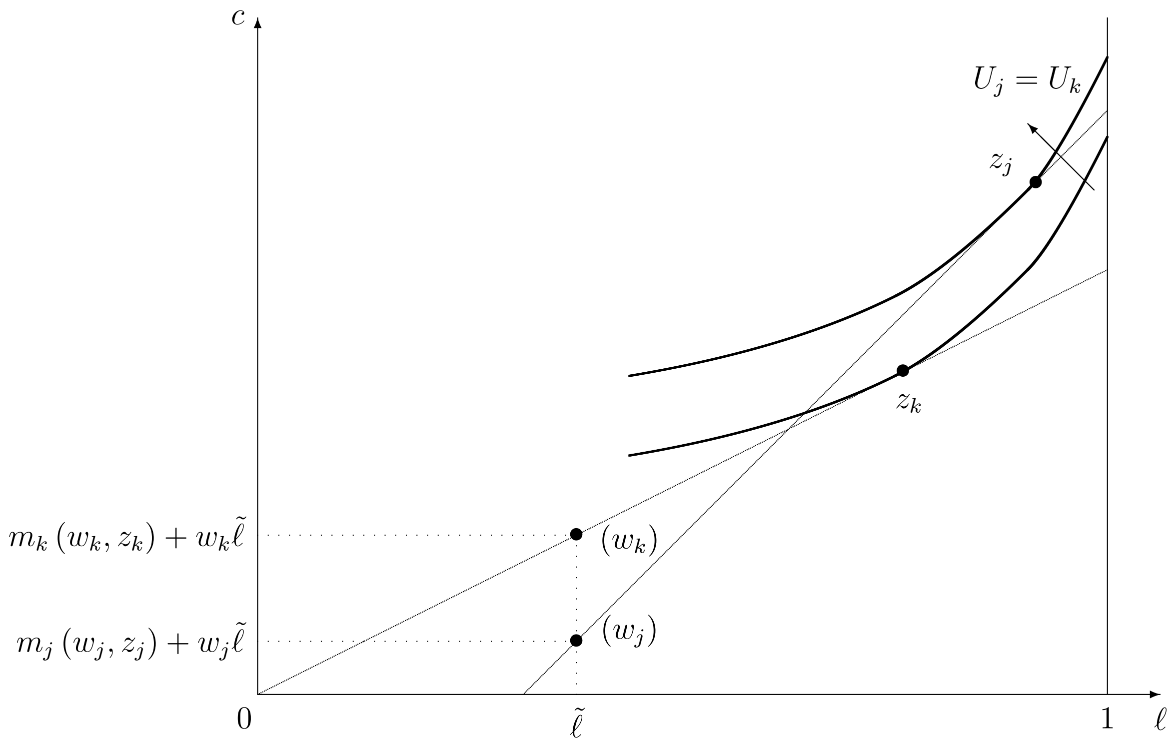

Figure 4: $U_j(z_j) > U_k(z_k)$ whereas $m_k(w_k, z_k) + w_k\tilde{\ell} > m_j(w_j, z_j) + w_j\tilde{\ell}$.

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*.

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 wage rates, 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 wage rates and give bundles on the same indifference curves to individuals with identical preferences.[30]

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 $\tilde{w}$.

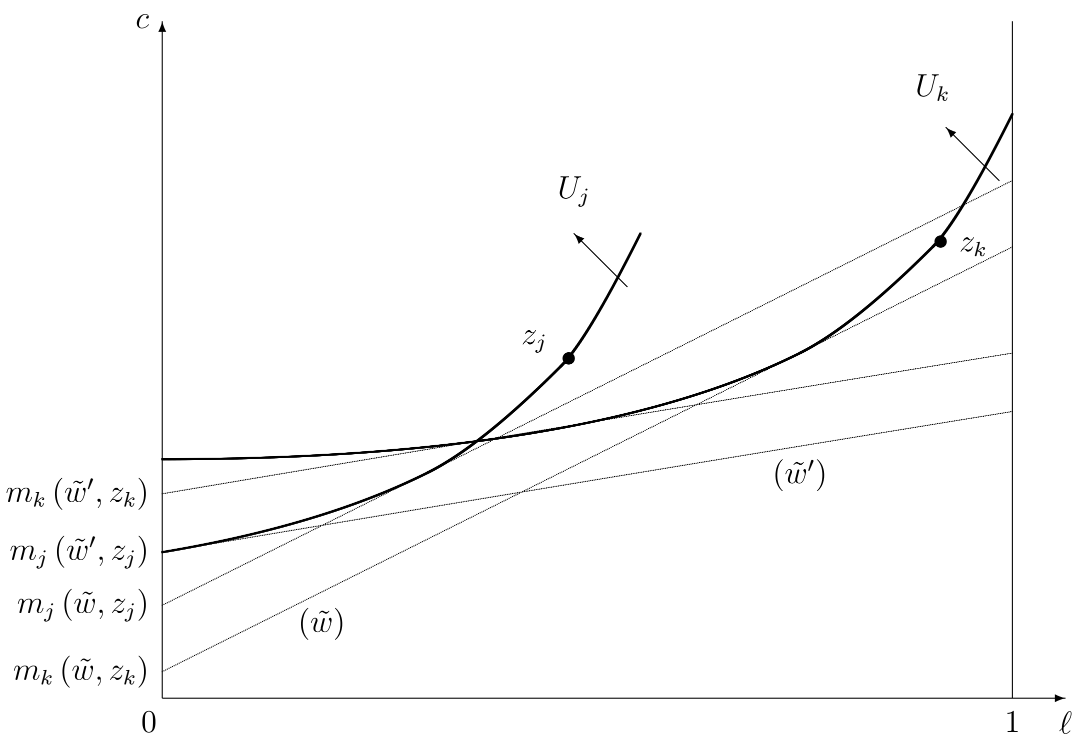

Figure 5: $m_j(\tilde{w}, z_j) > m_k(\tilde{w}, z_k)$ whereas $m_k(\tilde{w}', z_k) > m_j(\tilde{w}', z_j)$.

Two additional considerations suggest that the choice of $\tilde{w} = w_{\min}$ is

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

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 rate, 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).[31] 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.[32]

This section has shown that, even restricting attention to the class of indexes of the form

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

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 rate $w_{\min}$. 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 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: $U_i(\ell_i, c_i) = c_i - v\left(\frac{\ell_i}{\theta_i}\right)$, for some increasing and convex v function satisfying $v(0) = 0$. Let $\ell^*(\theta_i)$ be the optimal labor time of agent i if her wage rate 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

[31] See Fleurbaey and Maniquet (1996).

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

given). Then the index $m_i(\tilde{w}, z_i)$ boils down to

$$c_i - v\left(\frac{y_i}{w_i\theta_i}\right) - \tilde{w}\ell^*(\theta_i) + v\left(\frac{\ell^*(\theta_i)}{\theta_i}\right).$$

It is transparent that this utility index cannot be written as a function of the $w_i\theta_i$ products only. This means that optimal tax problems cannot be solved by the Brett and Weymark (2003) method in general.

9 Conclusion

In this paper, we have examined the contribution that notions of fairness can make to optimal 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, 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 the welfare economics baby with the utilitarian bath water, we have pleaded for going beyond the old 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 which 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 quite limited and cannot offer much because the weights cannot be transparently connected to fairness ideas. Even the method of weighting incomes directly, as proposed by Saez and Stantcheva (2013), cannot always be applied easily because even 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 hope that the last section 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 has been focused in this paper on the standard income tax model, and one could consider many possible enrichments of this model. Some of such enrichments (e.g., the addition of public goods, personal needs, or heterogenous households) would require rethinking the translation of fairness principles into utility indexes. The methodology laid out in this paper, hopefully, will be applicable to a wide set of contexts. Other taxation problems such as capital income taxation, commodity taxation, and Pigouvian taxation can be considered as well.[33]

Such extensions can also serve to address the worry that the preferences over labor and consumption that play an important role in the standard 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 the previous section, 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 rate 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. This is an exciting field for research.

References

[1] 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.

[2] Ambec, Stéphane and Lars Ehlers 2011, "Regulation via the Polluter-Pays Principle," Research Papers 2011-01, University of Montreal.

[3] Arneson, Richard J. 1989, "Equality of opportunity for welfare," Philosophical Studies 56, 77-93.

[4] Atkinson, Anthony B. 1995, Public Economics in Action, Oxford: Oxford University Press.

[33] See, e.g., Fleurbaey (2006) on commodity taxes and Ambec and Ehlers (2011) on externalities.

[5] Bergson (Burk), Abram 1938, "A reformulation of certain aspects of welfare economics," Quarterly Journal of Economics 52: 233–252.

[6] Besley, Timothy and Stephen Coate 1995, "The design of income maintenance programs," *Review of Economic Studies* 62: 187-221.

[7] Blackorby, Charles and David Donaldson 1988, "Money-metric utility: A harmless normalization?" Journal of Economic Theory 46(1) 120-129.

[8] Boadway, Robin 2012, From Optimal Tax Theory to Tax Policy, M.I.T. Press.

[9] Boadway, Robin, Maurice Marchand, Pierre Pestieau and Maria del Mar Racionero 2002, "Optimal redistribution with heterogeneous preferences for leisure," Journal of Public Economic Theory 4(4) 475-798.

[10] Brett, Craig and John A. Weymark (2003). "Financing education using optimal re-distributive taxation," Journal of Public Economics 87(11), 2549-2569.

[11] Choi, James J. , David Laibson, Brigitte C. Madrian and Andrew Metrick 2003, "Optimal defaults," American Economic Review 93: 180-185.

[12] Choné, Philippe, and Guy Laroque 2012, "Negative marginal tax rates and heterogeneity," American Economic Review 100(5): 2532-2547.

[13] Clark, Andrew E., Paul Frijters, Michael A. Shields 2008, "Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles," *Journal of Economic Literature* 46: 95-144.

[14] Clark, Andrew E. 2015, "Is happiness the best measure of well-being?," in M. Fleurbaey, M. Adler (eds.), Handbook of Well-Being and Public Policy, Oxford: Oxford University Press.

[15] Cohen, Gerald A. 1989 "On the currency of egalitarian justice" Ethics 99: 906-944.

[16] Deaton, Angus 1980, "The measurement of welfare: Theory and practical guidelines," *Living Standards Measurement Study Working Paper* No. 7, World Bank.

[17] Deaton, Angus, and John Muellbauer 1980, Economics and Consumer Behavior, Cambridge University Press.

[18] Diener, Ed, John F. Helliwell, Daniel Kahneman 2010, *International Differences in Well-Being*, Oxford: Oxford University Press.

[19] Di Tella, Rafael and Robert McCulloch 2006, "Some uses of happiness data in economics," Journal of Economic Perspectives 20: 25–46.

[20] 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: 71–85.

[21] Dworkin, Ronald 1981, "What is equality? Part 1: Equality of welfare," Philosophy and Public Affairs 10(3) 185-246.

[22] Dworkin, Ronald 1981, "What is equality? Part 2: Equality of resources," Philosophy and Public Affairs 10(4) 283-345.

[23] Fleurbaey, Marc, 1994, "On fair compensation" Theory and Decision 36, 277-307.

[24] Fleurbaey, Marc, 2006, "Is commodity taxation unfair?" Journal of Public Economics 90(10-11) 1765-87.

[25] Fleurbaey, Marc, 2008, Fairness, Responsibility and Welfare, Oxford University Press.

[26] Fleurbaey, Marc, 2010, "Assessing risky social situations" Journal of Political Economy 118(4) 649-680. 1765-87.

[27] Fleurbaey, Marc and François Maniquet 1996, "Fair allocation with unequal production skills : The no-envy approach to compensation", Mathematical Social Sciences 32: 71–93.

[28] Fleurbaey, Marc and François Maniquet 2005, "Fair social orderings with unequal production skills," Social Choice and Welfare 24: 1-35.

[29] Fleurbaey, Marc and François Maniquet 2006, "Fair income tax," Review of Economic Studies 73(1) 55-83.

[30] Fleurbaey, Marc and François Maniquet 2007, "Help the low-skilled, of let the hardworking thrive? A study of fairness in optimal income taxation," Journal of Public Economic Theory 9(3) 467-500.

[31] Fleurbaey, Marc and François Maniquet 2011a, *A Theory of Fairness and Social Welfare*, Cambridge University Press.

[32] Fleurbaey, Marc and François Maniquet 2011b, "Compensation and Responsibility,' in: K.J. Arrow, A.K. Sen and K. Suzumura (eds) *Handbook of Social Choice and Welfare*, Vol. 2, Amsterdam: North-Holland.

[33] Fleurbaey, Marc and François Maniquet 2011c, "The Kolm tax, the tax credit and the flat tax," in M. Fleurbaey, M. Salles, J. Weymark (eds.), *Social Ethics and Normative Economics. Essays in honor of Serge-Christophe Kolm*, Springer.

[34] Fleurbaey, Marc and François Maniquet 2012, *Equality of Opportunity: The Economics of Responsibility*, World Scientific Series in Economic Theory.

[35] Fleurbaey, Marc and Giacomo Valletta 2014, "Fair optimal tax with endogenous productivities," mimeo.

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

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

[38] Jacquet, Laurence and Etienne Lehmann 2014, "Optimal Nonlinear Income Taxation with Multidimensional Types: The Case with Heterogeneous Behavioral Responses," THEMA Working Paper n°2014-01.

[39] Jacquet, Laurence and Dirk Van de Gaer 2011, "A comparison of optimal tax policies when compensation and responsibility matter," Journal of Public Economics 95, 1248-1262.

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

[41] Kahneman Daniel and Alan B. Krueger 2006, "Developments in the measurement of subjective well-being," *Journal of Economic Perspectives* 20: 3–24.

[42] Kolm, Serge 1972, Justice et Equité, Paris: Editions du CNRS.

[43] Kaplow, Louis 2008, The Theory of Taxation and Public Economics, Princeton University Press.

[44] Kaplow, Louis and Steven Shavell 2001, "Any non-welfarist method of policy assessment violates the Pareto principle," Journal of Political Economy 109(2), 281-286.

[45] Konow, James 2001 "Fair and square: The four sides of distributive justice, " Journal of Economic Behavior & Organization 46(2), 137-164.

[46] Lockwood, Benjamin B. and Matthew Weinzierl 2014, "De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution," Harvard Business School WP 12-063.

[47] Mankiw, N. Gregory, Matthew Weinzierl and Canny Yagan 2009, "Optimal taxation in theory and practice," Journal of Economic Perspectives 23: 147-174.

[48] Mankiw, N. Gregory 2010, "Spreading the wealth around: reflections inspired by Joe the Plumber," Eastern Economic Journal 36: 285-298.

[49] Mirrlees, James 1971, "A exploration in the theory of optimum income taxation," Review of Economic Studies 38(2), 175-208.

[50] Mirrlees, James 1974, "Notes on welfare economics, information and uncertainty," in M. Balch, D. McFadden, and S. Wu (Eds.) Essays in Equilibrium Behavior Under Uncertainty, Amsterdam: North-Holland, 243-258.

[51] Mirrlees, James 1982, "The economic uses of utilitarianism," in A. Sen and B. Williams (Eds.) Utilitarianism and Beyond, Cambridge University Press.

[52] Negishi, Takashi 1972, General equilibrium theory and international trade, Amsterdam: North-Holland.

[53] Pazner, Elisha and David Schmeidler 1978, “Egalitarian-Equivalent Allocations: A New Concept of Economic Equity”, *Quarterly Journal of Economics* 92: 671-687.

[54] Piketty, Thomas, and Emmanuel Saez 2012 “Optimal labor income taxation,” NBER Working Paper 18521.

[55] Preston, Ian, and Ian Walker “Welfare measurement in labor supply models with nonlinear budget constraints,” Journal of Population Economics 12(3) 343-361.

[56] Rawls, John 1971, A Theory of Justice, Cambridge, Mass.: Harvard University Press.

[57] Rawls, John 1982, “Social unity and primary goods,” in A.K. Sen and B. Williams (eds.), Utilitarianism and Beyond, Cambridge: Cambridge University Press.

[58] Rochet, Jean-Charles, Lars Stole 2003, “The Economics of Multidimensional Screening,” in M. Dewatripont and L. Hansen (eds.), *Advances in Economic Theory. 7th World Congress*, Cambridge University Press, 2003.

[59] Robbins, Lionel 1937, *An Essay on the Nature and Significance of Economics*, London: Macmillan, 2nd revised ed.

[60] Roemer, John E. 1993, “A pragmatic theory of responsibility for the egalitarian planner,” Philosophy and Public Affairs 22(2), 146-166.

[61] Roemer, John E. 1996, “Harsanyi’s impartial observer is not a utilitarian,” in M. Salles and J. A. Weymark (eds) *Justice, Political Liberalism and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge: Cambridge University Press.

[62] Roemer, John E. 1998, Equality of Opportunity, Cambridge: Harvard University Press.

[63] Roemer, John E., Rolf Aaberge, Ugo Colombino, Johan Fritzell, Stephen P. Jenkins, Arnaud Lefranc, Ive Marx, Marianne Page, Evert Pommer, Javier Ruiz-Castillo, Maria Jesus San Segundo, Torben Tranaes, Alain Trannoy, Gert G Wagner and Ignacio Zubiri 2003, “To what extent

do fiscal regimes equalize opportunities for income acquisition among citizens," Journal of Public Economics 87(3-4), 539–565.

[64] Saez, Emmanuel 2001, "Using elasticities to derive optimal income tax rates," Review of Economic Studies 68(1) 205–229.

[65] Saez, Emmanuel 2002, "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," Quarterly Journal of Economics 117(3) 1039–1073.

[66] Saez, Emmanuel and Stephanie Stantcheva 2013, "Generalized marginal social welfare weights for optimal tax theory," NBER Working Paper No. 18835.

[67] Salanié, Bernard 2011, The Economics of Taxation, 2nd ed., MIT Press.

[68] Samuelson, Paul A. 1974 "Complementarity – An Essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," Journal of Economic Literature 12, 1255–1289.

[69] Sen, Amartya K. 1985, Commodities and Capabilities, New York: North-Holland.

[70] Sheffrin, Steven M. 2013, Tax Fairness and Folk Justice, New York: Cambridge University Press.

[71] Van Praag, Bernard M.S. and Ada Ferrer-i-Carbonell 2008, *Happiness Quantified*, 2nd rev. ed., New York: Oxford University Press.

[72] Varian, Hal 1974, "Efficiency, equity and envy", Journal of Economic Theory 9: 63–91.

[73] Vickrey, William 1945, "Measuring marginal utility by reactions to risk," Econometrica 13(4), 319–333.

[74] Weinzierl, Matthew 2012 "Why Do We Redistribute so Much but Tag so Little? Normative Diversity, Equal Sacrifice and Optimal Taxation." NBER Working Paper No. 18045.

[75] Wilson, Robert B. 1993, Nonlinear Pricing, New York: Oxford University Press.

[76] Yaari, Menahem and Maya Bar-Hillel 1984, "On dividing justly," Social Choice and Welfare 1, 1-24.