Fair income tax with endogenous productivities and a fresh start

Aitor Calo-Blanco[1] 📵

Abstract

This paper considers a model in which agents have heterogeneous preferences over labour and consumption. Additionally, they also differ in their earning skills, which are a function of both an innate ability and an early investment decision. In this framework we axiomatically derive a social ordering function that, besides compensating agents for their unequal productivities, grants a fresh start to those who regret their initial choices. Next, by assuming a second-best context we characterise the income tax scheme that satisfies this social ordering. This analysis permits us to present an explicit criterion for the assessment of social welfare under different tax policies. We obtain that the optimal scheme aims to compensate those endowed with the lowest marginal productivity. More importantly, on account of the forgiveness ideal positive social marginal weights are assigned to those who earn the lowest income levels, something that induces a progressivity tendency at the bottom of the earnings distribution.

Keywords Fairness · Endogenous skills · Forgiveness · Social preferences · Taxations

1 Introduction

In the last few years some papers have analysed the problem of implementing a fair redistribution that, while respecting heterogeneous individual preferences, it also aims to compensate agents for differences in their innate abilities. More recently, a different branch of the fairness literature has introduced the question of forgiveness, which refers to the possibility of compensating agents who have a change in their preferences, and who hence later regret their initial choices. Although these two ethical ideals, fair redistribution for unequal skills and forgiveness, have been separately studied in earlier works (e.g., Fleurbaey and Maniquet 2006; Valletta 2014; Fleurbaey 2005; Calo-Blanco 2016), the interaction between

---

✉ Aitor Calo-Blanco
  aitor.calo@udc.es

[1] Departamento de Economía, Universidade da Coruña, Campus de Elviña, 15071 A Coruña, Spain

both in a model in which individuals are endowed with heterogeneous traits has been barely analysed.[1]

Considering such an interaction provides new insights into the fairness literature because the combination of these two approaches in a model with endogenous productivities makes the construction of social preferences more complex. This is so because the unfair well-being losses that one agent suffers worsens when an early choice affects the individual marginal productivity, something that may additionally limit her consumption opportunities. Moreover, in second informational settings in which some personal traits are private information the interrelation between fair compensation and forgiveness, together with the assumption of endogenous productivities, amplifies the planner's difficulties of implementing any tax scheme. More precisely, the planner finds it more complicated to provide individuals with incentives to self-select the situation that it has designed for them, since they would have then more alternatives to replicate others' behaviour in order to get extra resources. Eventually, this reduces the extent of the fair redistribution policy.

Therefore, the first objective of the present paper is to construct, grounded on different normative principles, social preferences that will allow us to combine these two ethical ideals. We face this issue by considering a framework in which individuals have both unequal earning abilities and heterogeneous preferences over consumption and leisure. Moreover, we assume that individuals can affect their final productivity by means of an early investment choice, something that compounds the problem of regretting previous decisions because it affects the way in which individuals will transform labour into consumption. The second objective of this paper is to characterise a taxation scheme that implements such social preferences in a second-best context in which not all the pieces of information are available for the policy-maker.

As regards the literature on the above-mentioned topics, some of the most relevant theories of fairness and responsibility argue that individuals should only be compensated for outcome differences that are a result of elements for which they should not be deemed responsible (see Rawls 1971; Dworkin 1981a, 1981b; Arneson 1989; Cohen 1989; Roemer 1998). As a consequence of the contributions made by Fleurbaey and Maniquet (2011) and Fleurbaey and Maniquet (2018), a new branch of this fairness literature which focuses on unequal production skills and optimal taxation has emerged.[2] Unlike the traditional models, these authors deal with the compensation problem by resorting to fairness value judgments which consider solely individual non-comparable ordinal preferences. The first aim of this branch of the literature is to construct social preferences over labour-consumption allocations that, while respecting individual preferences, allow the planner to rank all scenarios when society is willing to pay compensations for differences in individual innate abilities (e.g., Fleurbaey and Maniquet 2005). Following Mirrlees' (1971) seminal contribution, the second goal of this new branch is to study the issue of social welfare and fair taxation. More precisely, under the assumption that only income levels are observable, this approach aims to provide criteria for the characterisation of a tax scheme that maximises social preferences which compensate individuals for their different productive skills, but not for their different preferences for working (e.g., Luttens and Ooghe 2007).

---

[1]An exception is Calo-Blanco (2014), who combines these two ethical views in a model in which individuals differ in both their health care needs and their preferences over health and consumption.

[2]Previous works dealing with this issue, although not from an axiomatical viewpoint, are provided by Bossert et al. (1999) and Schokkaert et al. (2004).

As we have previously mentioned, we focus on a recent extension of these models which introduces endogenous skills. For instance, Valletta (2014) and Fleurbaey and Valletta (2018) consider a scenario in which each agent is endowed with a marginal productivity of labour that depends on her investment in human capital (which is related to health and/or education). By assuming that individuals have heterogeneous preferences over consumption, labour and human capital, these authors provide some insight into the shape of the optimal tax scheme when individuals are partially responsible for their earning abilities. We extend this scarce literature by introducing the possibility of regretting the endogenous labour choices, something that drastically changes the fair compensation problem if one wants to provide individuals with a fresh start.

Interestingly enough, if the investment decision in one's own productivity is understood as an educational choice, this approach can be related to the literature on education and taxation. Examples of this literature are Bovenberg and Jacobs (2005) and Jacobs and Bovenberg (2011), who have studied how endogenous wages affect the design of education subsidies that aim to provide people with incentives to work harder. Since in our model the endogenous investment is a one-shot decision, it has some similarities to the (Findeisen and Sachs' 2016) approach. By including both uncertainty and heterogeneity in innate abilities, these authors analyse the problem of funding education when wages are determined by the costly education decisions made by the individuals before entering the labour market.[3] They defend that in a second-best scenario the use of student loans with income-contingent repayment rates allows the planner to minimise the efficiency cost of labour supply distortions. Although in a more simplified way and for different goals, our model also considers the possibility that the tax scheme distorts the education decision by stimulating some individuals to invest in one's own productivity. The difference between our paper and the (Findeisen and Sachs' 2016) approach is mainly twofold. On the one hand, apart from the skills heterogeneity we also assume that individuals differ in their preferences over labour and consumption. On the other hand, we endorse an environment without uncertainty in which individuals may regret *ex post* the investment decision in one's own productivity.

Consequently, besides the problem of compensating individuals for their unequal production skills, in the present paper we also endorse the forgiveness ideal which advocates that a fresh start should be conferred upon those agents who suffer a change in their preferences and who hence regret their previous labour decisions. As Fleurbaey (2005) and Fleurbaey (2008) specifies, these changes refer to genuine modifications in preferences which imply a firm determination to modify the lifestyle. That is, these changes are not associated with erroneous beliefs or differences in information about the true state of nature, but with the desire to have the current situation evaluated with a new set of preferences. This includes the fully rational agents who do not suffer from any myopia but who at some point decide to change their lifestyle, and hence regret the consequences of what they have done so far. When any of these agents reassesses her own situation with her new ambitions she suffers a considerable constraint in the achievement of her life goals.

Whether individuals should or should not be deemed responsible for such changes is a controversial question. Some are reluctant to back this view as they consider that it may entail both incentive and moral issues (e.g., Arneson 1989; Dworkin 2000; Dworkin 2002).

---

[3]A different approach to the same issue is presented by Stantcheva (2017), who studies the impact of second-best optimal taxation on the accumulation of human capital when individuals have heterogeneous abilities and this process of accumulation is risky. Unlike Findeisen and Sachs (2016), she focuses on the human capital accumulation over the life-cycle instead of on the college education choice.

Their argument is threefold. First, these authors claim that helping those who have misman-aged their resources generates incentive problems, as some individuals may fake regret in order to receive extra resources. Second, these authors also consider that it is unfair to help regretful individuals for a frugality they have never practised. Finally, it is also argued that this problem can be solved with an insurance against changes in preferences, and that those who fail to do so should suffer the consequences of such changes.

All these arguments against fresh starts have been thoroughly challenged by Fleurbaey (2005) and Fleurbaey (2008). According to him, in the absence of any cost to others basi-cally no-one would be against helping individuals who regret their past choices. Therefore, it seems that there are no moral problems associated with this ideal and consequently, as it happens with the majority of the redistribution policies, what should really matter is the balance between the gains obtained by those who benefit from the forgiveness ideal and the losses experienced by those who have to fund it. Hence, the real potential problem that this ideal generates is that some individuals may strategically misrepresent their preferences in order to get additional resources, leading, this way, to a possible reduction of the extent of the redistribution policy. Nevertheless, Fleurbaey (2005) defends that a properly designed incentive-compatible fresh start policy limits any 'undeserved' compensation that individ-uals may receive.[4] Finally, this author also rejects the solution of giving the population the possibility of taking an insurance against changes in preferences. Apart from the intrinsic failures of insurance markets regarding *ex ante* and *ex post* evaluations, Fleurbaey (2008) argues that the insurance is not feasible in the case of forgiveness. Since regret is not observ-able, all those who have taken the insurance would eventually apply for the compensation regardless of what had happened with their preferences. Consequently, the insurance mar-ket would collapse unless the premium would be equal to the compensation, something that would make the insurance useless.

As a result of the previous discussion, our first aim in this paper is to construct a social ordering function that will allow us to rank all possible labour-consumption allocations when individuals are endowed with unequal preferences and production skills. This pro-ductivity is a function of an early investment choice that every agent has to make prior to obtaining her labour-consumption bundle. Unlike previous results, the social ordering that we derive is grounded on efficiency, robustness and fairness principles which satisfy responsibility and forgiveness criteria. Consequently, it aims to compensate both those who are poorly endowed and those who genuinely regret their labour decisions. The resulting social preferences give absolute priority to that individual associated with the lowest value of a comparable measure of well-being that entails neither skill inequalities nor regret in the individual labour decisions.

Once having characterised the social preferences, the next natural step to take is to evaluate the implementation of these preferences by means of a tax scheme. Despite the renewed interest in the topic of fairness and taxation that has emerged lately, there is vir-tually no study that analyses fair optimal taxation with forgiveness. The only exception is Calo-Blanco (2017), who proposes, in a health model, a scheme of taxes and treatments that maximises social preferences when those who regret their initial lifestyle decisions are granted a fresh start. Nevertheless, this author exclusively focuses on a framework in which health care needs are fixed and exogenously given. Therefore, the second aim of the present

---

[4]Real-life examples of a fresh start policy are the cases in which society helps those who later in their life want to finish their studies after having dropped out of school, or when a public health service treats all individuals who are in a bad health condition, regardless of their previous lifestyle.

paper is to characterise a fresh start taxation policy that maximises social welfare according to the obtained social preferences. It is important to clarify that this sort of preferences are designed to produce a full and optimal ranking of all the social allocations, and hence they are traditionally constructed in first-best contexts in which all pieces of information can be observed by the planner. However, when introducing taxation into the model it is usually assumed that some factors, such as marginal productivities and labour choices, are private information of the agents (e.g., Fleurbaey and Maniquet 2011). Consequently, we have to focus our taxation analysis on a second-best scenario in which the tax scheme has to be defined by means of monetary transfers that exclusively depend on the individual pre-tax income (see Mirrlees 1971). This characterisation is a complex task since the incentive problems that the informational constraints cause in the design of the optimal tax scheme are considerably exacerbated by the inclusion of both endogenous productivities and forgiveness. Despite all this we are able to describe the distribution of the pre-tax income and final consumption that the incentive-compatible tax scheme generates. We derive a simple formula for the social evaluation of different income taxation schemes, and we obtain that the optimal policy should give top priority to those who suffer the burden of a low earning ability. Moreover, opposite to previous results regarding fair taxation we prove that the policy yields an increasing marginal tax rate for those at the bottom of the income distribution. This outcome, which shows the effect of implementing the forgiveness principle, implies that society is willing to pay additional compensations to those who regret being at the very end of the income distribution.

The rest of the paper is organised as follows. Section 2 presents the basic components of the model. Section 3 introduces the ethical requirements that society endorses, while Section 4 presents the derivation of the social preferences. Section 5 develops the analysis of the optimal taxation in a second-best context, and it characterises the fresh start policy that maximises social preferences. Section 6 offers the conclusions of this study. All the proofs are gathered in the Appendices A and B.

2 The model

Our framework follows the fair social choice approach to taxation developed by Fleurbaey and Maniquet (2006) and Fleurbaey and Maniquet (2011). Let us consider a society in which there are only two goods; namely, labour, $\ell \in [0, 1]$, and consumption, $c \in \mathbb{R}_+$. The population in this society consists of a finite set of individuals $N = \{1, \ldots, i, \ldots, n\}$. Every agent $i \in N$ has a *labour-consumption bundle* $z_i = (\ell_i, c_i) \in Z = [0, 1] \times \mathbb{R}_+$ that designates the situation in which she has labour time ℓ_i and a level of consumption c_i. An *allocation* describes all the individuals' bundles, that is, $z_N = (z_i)_{i \in N} \in Z^n$.[5]

Any individual's earning ability is defined by her marginal productivity, or *innate skill*, and it is measured in consumption units. As we have mentioned in the Introduction, prior to deciding the amount of labour time each individual can partially modify her productivity. More precisely, any agent i's skill is a mapping s_i^α that depends on whether she has invested ($\alpha = 1$) or not ($\alpha = 0$) in improving her innate ability, and which results in a fixed parameter with a value of either $s_i^0 \in \mathbb{R}_{++}$ or $s_i^1 \in \mathbb{R}_{++}$, where $s_i^0 < s_i^1$. This initial investment entails, for all $i \in N$, an expenditure $\epsilon > 0$ if $\alpha = 1$, and 0 otherwise. Additionally, we

---

[5]A group of objects $a_N = (a_i)_{i \in N}$ denotes a list such as $(a_1, \ldots, a_i, \ldots, a_n)$.

consider that $s_i^1 - \epsilon > s_i^0$, and let $\widetilde{\ell_i} \in (0, 1)$ be such that $s_i^0 \widetilde{\ell_i} = s_i^1 \widetilde{\ell_i} - \epsilon$.[6] Let $\mathcal{S}$ be the set of all these mappings, in which $s_L^\alpha, s_H^\alpha \in \mathcal{S}$ determine, respectively, the lowest and the highest ones, that is, for all $i \in N$ and $\alpha = \{0, 1\}$ we have $s_L^\alpha \leq s_i^\alpha \leq s_H^\alpha$. Such limits are considered to be fixed for all possible allocations. Accordingly, for any $j, k \in N$ and $\alpha = \{0, 1\}$ individual j is said to have a better marginal productivity than k if and only if $s_k^\alpha \leq s_j^\alpha$. Let the population's profile of the abilities in this society be $s_N^\alpha = (s_i^\alpha)_{i \in N} \in \mathcal{S}^n$. This trait, together with the amount of labour time ℓ_i and her investment choice $\alpha_i = \{0, 1\}$, enables agent $i \in N$ to obtain pre-tax earnings $s_i^{\alpha_i} \ell_i - \epsilon \alpha_i \in \mathbb{R}_+$.

Every agent $i \in N$ has well-defined preferences R_i over the labour-consumption space Z, which are described by a complete preorder, that is to say, a binary relation that is reflexive, transitive and complete. The preferences, apart from being a complete preorder, must also be continuous, strictly convex and strictly monotonic. Let $\mathcal{R}$ denote the set of such preferences, and let the expression $z_i R_i z_i'$ denote that individual i weakly prefers z_i to bundle z_i'. The corresponding strict preference and indifference are denoted by P_i and I_i respectively. A profile of preferences in this society is denoted by $R_N = (R_i)_{i \in N} \in \mathcal{R}^n$.

In order to introduce the possibility of regret in the model, let us consider that agents make their choices according to some *ex ante* preferences $R_N^a = \left(R_i^a\right)_{i \in N} \in \mathcal{R}^n$, although they get their final utility from an *ex post* profile $R_N^p = \left(R_i^p\right)_{i \in N} \in \mathcal{R}^n$ that may or may not coincide with their *ex ante* preferences.

An *economy* is then described by a list $e = (s_N^\alpha, R_N^a, R_N^p) \in \mathcal{E}$. An allocation $z_N \in Z^n$ is said to be feasible for any economy $e \in \mathcal{E}$ if:

$$\sum_{i \in N} c_i \leq \sum_{i \in N} \left(s_i^{\alpha_i} \ell_i - \epsilon \alpha_i\right).$$

A *social ordering function*, SOF hereafter, **R** maps every element in this domain to a social ordering which defines a complete preorder over all the (feasible and not feasible) allocations in terms of forgiveness and fair redistribution. For any $e \in \mathcal{E}$ and $z_N, z_N' \in Z^n$, we write $z_N \mathbf{R}(e) z_N'$ to denote that allocation z_N is at least as good as z_N'. $z_N \mathbf{P}(e) z_N'$ means that z_N is strictly better than z_N', and $z_N \mathbf{I}(e) z_N'$ that they are equivalent.

It is important to mention that in our economy the *ex ante* preferences are the deciding factor in the determination of the allocations that are going to be evaluated by society. However, this evaluation is going to be implemented with individual *ex post* preferences because it is considered that only final goals matter (see Fleurbaey 2005; Calo-Blanco 2016). There are some good reasons to adopt this standpoint. The social degree of forgiveness is basically a normative choice, and hence it should be decided by society, similar to what happens with other normative decisions. But if we are able to build a model in which it is possible to compensate individuals for their maximum level of regret, any other intermediate standpoint selected by that society would also be feasible. That is, it would be possible to construct any social ordering in which the final result would be defined by the way in which the society wants to combine the *ex ante* and *ex post* preferences.

Once having introduced the basic components of the model, let us now present the concept of *labour-consumption budget set*. This concept, which is key at the time of constructing our social ordering function, describes the bundles that are feasible for any agent $i \in N$ after receiving a given lump-sum transfer t_i. Formally,

---

[6]This endogenous skill bears some resemblance to the one proposed by Valletta (2014). Alternatively, it can also be interpreted as the sort of learning effect in which, with no investment whatsoever, the marginal productivity increases in the quantity of labour.

Definition 1 For all $i \in N$, $s_i^\alpha \in \mathcal{S}$ and some $t_i \in \mathbb{R}$, the individual i's labour-consumption budget set is:

$$B(t_i, s_i^\alpha) = \left\{ (\ell, c) \in Z \mid \{c \leq s_i^0 \ell + t_i\} \cup \{c \leq s_i^1 \ell - \epsilon + t_i\} \right\}.$$

The graphical illustration of the framework that we have just introduced in this section is provided in Fig. 1. The solid line in the left-hand part of the picture delimits the individual labour-consumption budget set. For any agent $i \in N$ the feasible bundles are initially defined by her innate skill, which characterises a budget set with two parts (the different slopes are in parentheses below the lines). As we can observe in Fig. 1a, for low values of ℓ the individual does not find it optimal to invest in increasing her own ability. For instance, when $\ell = 0$ the difference between investing and not investing in her own productivity (depicted by the vertical distance between the solid and dashed straight lines) equals ϵ. By contrast, when the level of labour is sufficiently large, that is $\ell \geq \tilde{\ell}_i$, the gains associated with a larger skill outweigh the cost of investing in one's own productivity. Consequently, the budget line pivots upwards at this specific point, showing that the individual becomes more productive. Apart from this endogenous productivity, the other factor that characterises this labour-consumption budget set is the existence of a lump-sum transfer t_i, which moves this set upwards or downwards depending on the sign of the transfer. In Fig. 1a the set starts from above the origin, which means that individual i is receiving a positive transfer, something that allows her to have access to more bundles than in the *laissez-faire* scenario.

Figure 1b represents the general idea behind the forgiveness framework that we have assumed. Let us consider an individual $i \in N$ who is endowed with a production skill $s_i^\alpha \in \mathcal{S}$ which defines her feasible set of alternatives as a function of her investment choice (solid black line). Let us now assume that agent i makes her initial choice from this set according to the *ex ante* preferences $R_i^a \in \mathcal{R}$ (black curve), something that results in a labour-consumption bundle $z_i \in Z$. However, after having made this choice she wishes she had used alternative preferences $R_i^p \in \mathcal{R}$ (gray curves), and therefore she regrets her initial decision. As we can observe in Fig. 1b, this agent ends up with a level of utility that is lower than the one she could have obtained if she had made her initial choice with her *ex post* preferences R_i^p, that is $\hat{z}_i \in Z$.
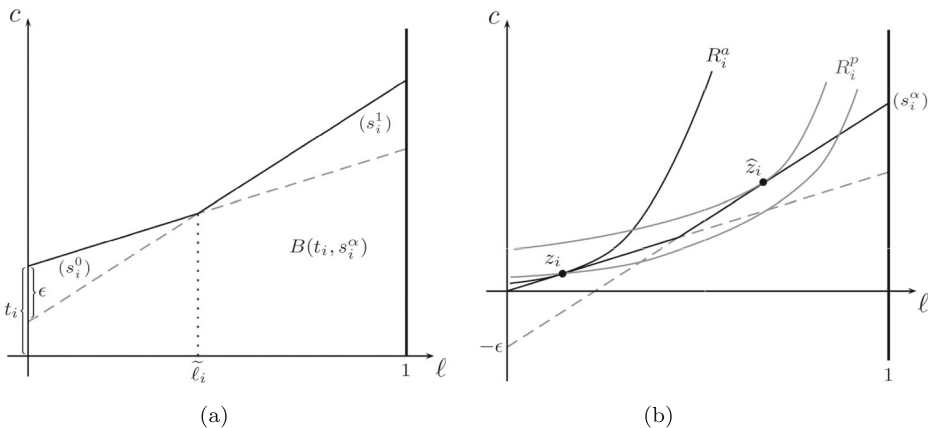

Fig. 1 Labour-consumption set and individual preferences

3 The ethical principles

Based on the definitions that we have introduced in the previous section, let us now present the ethical principles that are desirable for our social ordering function. In accordance with the forgiveness view which considers that changes in preferences are a legitimate cause for compensation, these principles are defined in terms of the *ex post* preferences.

The first axiom is the Pareto condition, which is a minimal requirement that ensures that the solution is efficient:

Axiom 1 (Strong Pareto) *For all $e \in \mathcal{E}$ and $z_N, z'_N \in Z^n$, if $z_i R^p_i z'_i$ for all $i \in N$; then $z_N \mathbf{R}(e) z'_N$. If moreover, $z_j P^p_j z'_j$ for some $j \in N$; then $z_N \mathbf{P}(e) z'_N$.*

The second ethical principle is a robustness requirement demanding that indifferent agents should not influence social preferences. More precisely, we adopt a version of the well-known separability property (see D'Aspremont and Gevers 1977) which entails that adding or removing agents whose bundles remain unchanged between two different allocations does not modify the social ordering over such allocations.

Axiom 2 (Separation) *For all $e \in \mathcal{E}$, $z_N, z'_N \in Z^n$ and $G \subset N$ such that $z_i = z'_i$ for all $i \in N \setminus G$, then:*

$$z_N \mathbf{R}(e) z'_N \Leftrightarrow z_G \mathbf{R}(e_G) z'_G,$$

where z_G and e_G are, respectively, the allocation and the economy associated with the reduced population G.

Moreover, and in line with the spirit of Arrow's 1951 condition of independence of irrelevant alternatives, we limit the pieces of information about the individuals' preferences that are required to compare two different social situations. More precisely, we demand social preferences over two allocations to depend only on the *ex post* indifference curves at these mentioned pair of allocations (e.g., Hansson 1973; Pazner 1979; Fleurbaey and Maniquet 2011).

Axiom 3 (Independence) *For all $e = (s^\alpha_N, R^a_N, R^p_N)$, $e' = (s^\alpha_N, R^{a'}_N, R^{p'}_N) \in \mathcal{E}$ and $z_N, z'_N \in Z^n$, if for all $i \in N$ and $z'' \in Z$:*

$$z_i I^p_i z'' \Leftrightarrow z_i I^{p'}_i z'',$$
$$z'_i I^p_i z'' \Leftrightarrow z'_i I^{p'}_i z'';$$

then $z_N \mathbf{R}(e) z'_N \Leftrightarrow z_N \mathbf{R}(e') z'_N$.

In order to get our characterisation results we need to combine these three efficiency and robustness requirements with principles that model redistributions. These principles should aim to increase social welfare by means of reducing well-being inequalities between individuals.

The first redistributive axiom that we consider endorses the ethical view that states that individuals who only differ in their production skills should obtain similar outcomes. More precisely, it establishes that a reduction in the consumption inequality between two agents with the same *ex post* preferences and labour time increases social welfare. Additionally, we consider that if these two individuals swap their bundles social preferences do not change, that is, agents are treated anonymously. Specifically:

Axiom 4 (Equal Preferences Transfer) *For all* $e \in \mathcal{E}$ *and* $z_N, z'_N \in Z^n$, *if there exist* $j, k \in N$ *and* $\delta \in \mathbb{R}_{++}$ *such that* $R^p_j = R^p_k$, $\ell_j = \ell_k = \ell'_j = \ell'_k$,

$$c'_j - \delta = c_j > c_k = c'_k + \delta,$$

and $z_i = z'_i$ *for all* $i \neq j, k$; *then* $z_N \mathbf{R}(e) z'_N$. *If otherwise* $c_j = c'_k$ *and* $c'_j = c_k$, *then* $z_N \mathbf{I}(e) z'_N$.

To make welfare comparisons that include agents who differ in their *ex post* preferences we have to introduce a second redistributive requirement. This additional principle endorses the ethical view that states that individuals should only be held responsible for their true (final) goals. Hence, those with equal abilities should be free to choose, with the *ex post* preferences, their most preferred bundle in an identical opportunity set. However, as Fleurbaey and Maniquet (2005) have thoroughly shown, this responsibility requirement generally clashes with the previous compensation principle which aims to reduce inequalities between individuals with identical preferences. To accommodate these two redistributive requirements we opt to weaken our responsibility axiom by restricting it only to agents who are endowed with a specific reference production skill, more precisely, to those who have the lowest innate ability. Consequently, if all agents have this particular skill and they do not regret their prior choices, the optimal policy should be the *laissez-faire* allocation with no redistribution whatsoever. This is so because in such a scenario differences in personal outcomes would exclusively reflect different autonomous decisions, including the fixed initial choice of one's own productivity. Finally, the axiom also establishes that similar individuals have to be treated anonymously. Formally,

Axiom 5 (s^α_L-Skill Transfer) *For all* $e \in \mathcal{E}$ *and* $z_N, z'_N \in Z^n$, *if there exist* $j, k \in N$ *and* $\delta \in \mathbb{R}_{++}$ *such that* $s^\alpha_j = s^\alpha_k = s^\alpha_L$,

$$z_j \in \max |_{R^p_j} B(t_j, s^\alpha_j), z'_j \in \max |_{R^p_j} B(t'_j, s^\alpha_j),$$

$$z_k \in \max |_{R^p_k} B(t_k, s^\alpha_k), z'_k \in \max |_{R^p_k} B(t'_k, s^\alpha_k),$$

$$t'_j - \delta = t_j > t_k = t'_k + \delta,$$

and $z_i = z'_i$ *for all* $i \neq j, k$; *then* $z_N \mathbf{R}(e) z'_N$. *If otherwise* $t_j = t'_k$ *and* $t'_j = t_k$, *then* $z_N \mathbf{I}(e) z'_N$.

Axiom 5 establishes that between two allocations that differ only in the situation of two non-regretful agents endowed with s^α_L, and which may entail different initial investment choices for them, society ranks first the one in which the budget inequality between them is the smallest. The choice of s^α_L entails some interesting features that need to be mentioned (see Fleurbaey and Maniquet 2018). First, with this reference skill agents who are more labour-averse will be favoured by social preferences. Hence, such a low value may be a good choice if one considers that work aversion is partly due to low job quality for the unskilled. Second, s^α_L is the only value in the population's profile of abilities that guarantees that redistribution will never violate the participation constraint. Finally, the lowest innate ability emerges as the reference skill when one specific normative principle is assumed. This principle establishes that a redistribution from a rich individual to a poor one, both endowed with the same skill, enhances welfare provided that the former has a lower aversion to labour than the latter, and also that the consumption inequality between both is larger than what the difference in labor time would justify.

The s_L^α-Skill Transfer is illustrated in Fig. 2 for an economy with two individuals, j and k, and a given allocation $z_N' \in Z^2$. As required by the principle, both agents are maximising their *ex post* preferences endowed with the lowest innate ability. According to Axiom 5, an alternative allocation such as $z_N \in Z^2$ is socially superior because it entails a smaller budget inequality between the agents. Note that in this particular example the *ex ante* investment choices differ. This does not mean that individuals can change their early decisions, but that our society is able to rank any two pair of social allocations that satisfy the conditions required by the axiom.

It is important to stress that the way in which these ethical principles are defined permits us to tackle both differences in production skills and the possibility of regret. Hence, the axioms aim to compensate individuals for the two aspects that our study is concerned about.

4 Social preferences

After having presented the ethical principles that our society endorses, in this section we proceed to define the social preferences that these principles induce. In order to fulfill this analysis let us first introduce the following concept:

Definition 2 For all $e \in \mathcal{E}$ and $z_i \in Z$, the individual i's s_L^α-Implicit Transfer is the scalar $T_i(z_i, R_i^p) \in \mathbb{R}$ that satisfies:

$$T_i(z_i, R_i^p) = t \Leftrightarrow z_i \, I_i^p \max |_{R_i^p} B(t, s_L^\alpha).$$

The s_L^α-*Implicit Transfer* is a comparable measure of individual well-being that relates any agent's *ex post* evaluation of her own situation to a specific money-metric utility. This metric is defined as the minimum lump-sum transfer that leaves the individual *ex post*
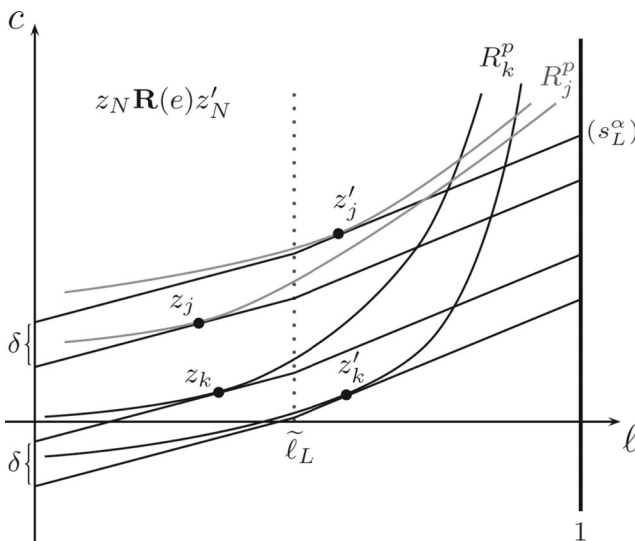
Fig. 2 The s_L^α-Skill Transfer

indifferent between her current bundle and being free to choose her labour time according to R_i^p in a budget set defined by this transfer and the lowest earning ability.

A graphical representation of this money-metric is provided in Fig. 3a, in which individual $i \in N$, who is endowed with a production skill s_H^α, chooses bundle $z_i \in Z$ with her *ex ante* preferences $R_i^a \in \mathcal{R}$. Let us now consider that, *ex post*, she changes her preferences and cares more about labour. In order to evaluate this agent's well-being loss, we can observe that endowed with s_L^α she could have obtained exactly the same *ex post* utility after receiving a negative transfer equal to $T_i(z_i, R_i^p)$ (see Fig. 3a). We can also see how the endogenous productivity deepens the well-being loss that this regret causes. For instance, if the reference ability would be fixed and equal to s_L^0 the well-being loss would be smaller than the actual one (see the dashed straight line in the same picture). That is, since the individual can increase her own skill, something that is optimal when she shows a low disutility of labour, the reduced productivity associated with a regretted small choice of labour may generate a further utility loss. Besides evaluating the cost of mismanaging the choice of labour, the s_L^α-Implicit Transfer also assesses the disadvantage related to a poor endowment. For instance, if the agent had made the same mistake but endowed with a strictly lower marginal productivity than s_H^α, the social evaluation of her well-being would have been smaller than the one depicted in Fig. 3a, showing, this way, the additional loss due to a lower innate ability.

Let us now make use of the concept of s_L^α-Implicit Transfer to define the following SOF:

Social Ordering Function 1 (s_L^α-Implicit Leximin) *For all $e \in \mathcal{E}$ and $z_N, z'_N \in Z^n$,*

$$z_N \mathbf{R}_{s_L^\alpha}^{\text{lex}}(e) z'_N \Leftrightarrow (T_i(z_i, R_i^p))_{i \in N} \geq_{\text{lex}} (T_i(z'_i, R_i^p))_{i \in N}.$$

This social ordering ranks first that allocation in which the lowest value, in lexicographic terms, of the s_L^α-Implicit Transfer across the population is the highest. Let us now present the scenario in which $\mathbf{R}_{s_L^\alpha}^{\text{lex}}$ is obtained as the social preferences.

Theorem 1 *On the domain $\mathcal{E}$, a social ordering function satisfying Strong Pareto, Separation, Independence, Equal Preferences Transfer and s_L^α-Skill Transfer is a s_L^α-Implicit Leximin function.*
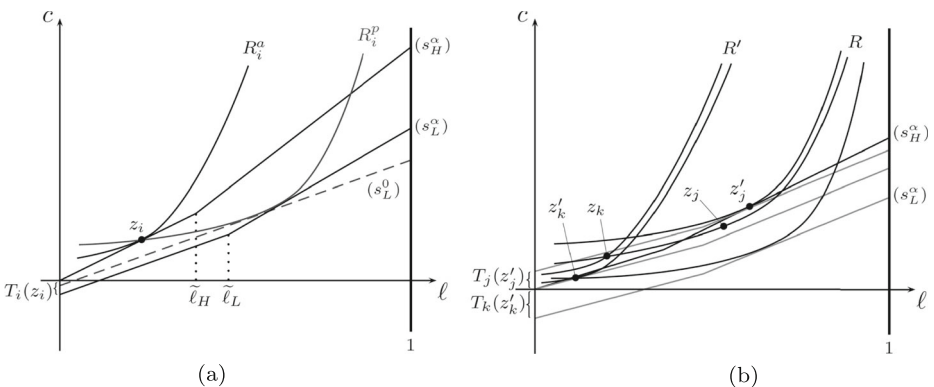

Fig. 3 Individual well-being and social preferences

Theorem 1 characterises the s_L^α-Implicit Transfer function as the right measure to make social welfare assessments when individuals who are endowed with different endogenous marginal productivities may regret their initial labour choices. Figure 3b illustrates this social ordering function.

Let us consider an economy with only two individuals; namely, j and k. Agent j both has preferences $R \in \mathcal{R}$ with a low disutility of work and is endowed with the largest production skill. By contrast, k exhibits preferences $R' \in \mathcal{R}$ with a higher labour disutility, and moreover her innate ability is the lowest one in the economy. In such a scenario the *laissez-faire* allocation is characterised by $z_N' \in Z^2$ (see Fig. 3b). Additionally, let us assume that after having made her choice individual k changes her preferences to the set R. As we can observe in the picture, our comparable measure of individual well-being yields a positive value for j and a negative one for k. That is, each individual is indifferent between remaining in her own *ex post* situation and being free to choose her labour time, with her final preferences, in a budget set defined by s_L^α and the lump-sum transfers $T_j(z_j', R) > 0$ and $T_k(z_k', R) < 0$, respectively. According to the s_L^α-Implicit Leximin SOF, any policy that induces an alternative allocation such as $z_N \in Z^2$ improves social welfare, that is $z_N \mathbf{P}_{s_L^\alpha}^{\text{lex}}(e)z_N'$. This is so because the minimum s_L^α-Implicit Transfer in this second allocation is higher than the smallest value in z_N', as we can observe in Fig. 3b. The task of describing how to induce agents to select this allocation z_N by means of a fresh start tax policy will be undertaken in the next section.

5 Optimal taxation with a fresh start

So far we have established a specific ranking that allows us to make welfare assessments when society wants both to compensate individuals for differences in their production skills and to give those who genuinely regret their choice a fresh start. In the present section we turn our analysis to the study of an incentive-compatible policy that puts such a ranking into practice. For ease of exposition, let us introduce the following simplifications.

First, we assume that individual preferences satisfy the *single-crossing property*. This implies that any two indifference curves of two different preferences cross no more than once. More precisely, for any $(\ell, c), (\ell', c') \in Z$ it is said that individual preferences $R_j \in \mathcal{R}$ present a lower aversion to labour than those of $R_k \in \mathcal{R}$, something we denote by $R_j \succ^\ell R_k$, if they satisfy the following relations:

$$\begin{cases} \ell > \ell' \text{ and } (\ell, c)I_k(\ell', c') \Rightarrow (\ell, c)P_j(\ell', c') \\ \ell < \ell' \text{ and } (\ell, c)I_j(\ell', c') \Rightarrow (\ell, c)P_k(\ell', c'). \end{cases}$$

Second, let us assume that any individual who *ex post* changes her preferences becomes less averse to labour, that is, for all $i \in N$ either $R_i^p = R_i^a$ or $R_i^p \succ^\ell R_i^a$. Additionally, we consider that there exists a finite number of types of individual preferences $\mathcal{R} = \{R_1, \ldots, R_f, \ldots, R_F\}$, which are ranked according to their aversion to labour, with $R_{f+1} \succ^\ell R_f$ for all $R_f \in \mathcal{R} \setminus \{R_F\}$. The reason why we assume that agents never become more averse to labour is because they can *ex ante* affect their own productivity. If one considers this investment decision as an educational choice, Fleurbaey (2008) defends that in this particular case the regret should go only in one direction, with people regretting having stopped too early.

Third, let us consider that there are only two types of innate abilities, and hence individuals have either a *high* or a *low* skill. Therefore, let $\mathcal{S} = \{s_L^\alpha, s_H^\alpha\}$ be the set of the

possible marginal productivities, with $s_L^\alpha < s_H^\alpha$ for all $\alpha = \{0, 1\}$. Moreover, we assume that $s_H^0 = s_L^1 - \epsilon$ and that $s_H^0 \tilde{\ell}_H = s_L^0 \tilde{\ell}_L = \tilde{y}$.

Finally, for any given type of preferences $R_f \in \mathcal{R}$ and any given ability $s^\alpha \in \mathcal{S}$ there exist individuals who stick to this particular given type, and individuals who would change to it from any other type of preferences that exhibit a higher aversion to labour.

Let us now introduce a social planner that aims to maximise the lowest value of the s_L^α-Implicit Transfer across individuals, as we have established in Theorem 1. As it is standard in this sort of taxation models (e.g., Mirrlees 1971; Fleurbaey and Maniquet 2011) we consider a second-best context in which only pre-tax earned income $y_i = s_i^{\alpha_i} \ell_i - \epsilon \alpha_i$ is observable. Given this informational constraint, the planner's fresh start policy is defined as a tax scheme that characterises monetary transfers depending exclusively on this variable, that is $\tau(y) : \mathbb{R}_+ \to \mathbb{R}$. The tax turns into a subsidy when $\tau(y) < 0$. This scheme is used by the planner to design the optimal redistribution policy by means of changing any individual's budget set, and hence her initial pre-tax income choice. More precisely, in the second-best scenario any agent i's final consumption is now determined by her net or post-tax income:

$$c_i \leq y_i - \tau(y_i).$$

According to this informational context, any second-best optimal allocation has to satisfy the following two conditions. First, it has to be *feasible*, that is, $\sum_{i \in N} \tau(y_i) \geq 0$. This implies that the total consumption in the final allocation must not exceed the aggregate earnings. Second, since individuals are free to choose their pre-tax income in the budget set modified by the tax function, the allocation induced by this function must be *incentive-compatible*. This means that no individual *ex ante* prefers the bundle of any other agent, provided that such a bundle is feasible for her. To formally define this condition in our informational structure it is convenient to focus on the earnings-consumption space, in which any individual i's bundle is described by $x_i = (y_i, c_i) \in X = \mathbb{R}_+^2$. For every individual $i \in N$ who is endowed with a production skill $s_i^\alpha \in \mathcal{S}$, let $\widehat{R}_i^{s_i}$ denote her preferences over such bundles, which can be derived from the ordinary preferences $R_i \in \mathcal{R}$ as follows (see Fleurbaey and Maniquet 2011):

$$(y, c)\widehat{R}_i^{s_i}(y', c') \Leftrightarrow \left(\frac{y + \epsilon\alpha_i}{s_i^{\alpha_i}}, c\right) R_i \left(\frac{y' + \epsilon\alpha_i'}{s_i^{\alpha_i'}}, c'\right),$$

where α_i (respectively α_i') is the investment choice that agent i makes to get income y (respectively y').

Figure 4 depicts this second-best scenario. As we can observe in the right-hand side of the picture, and contrary to the case in which labour is observable (left-hand side), the slope of the adapted *laissez-faire* budget set is equal to 1 for all agents. This slope represents the one-to-one trade-off between pre-tax income and consumption in the absence of any tax or transfer. The boundary of this new set is no longer delimited by $\ell = 1$, but by the maximum earnings that one individual can obtain.

In this scenario any individual's optimisation problem consists of choosing the level of income, within a range that is feasible for her, that maximises her preferences. For instance, let us consider an economy with two individuals, $j, k \in N$, who have the same preferences over labour and consumption $R \in \mathcal{R}$, but who are endowed with different innate abilities, more precisely $s_j^\alpha = s_H^\alpha$ and $s_k^\alpha = s_L^\alpha$. Let us assume that both are indifferent between investing in productivity and not doing it (see the left-hand side of Fig. 4). The right-hand side of the picture reproduces this initial situation in the space X, in which every agent is characterised by two curves that represent the same level of utility as a function of the
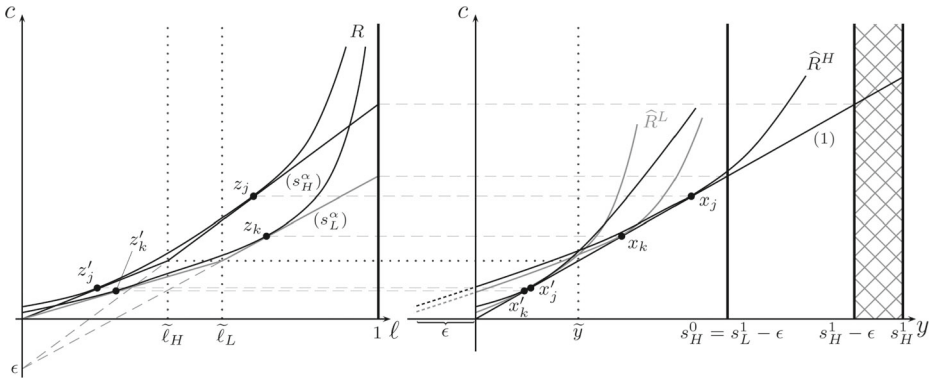
Fig. 4 *Laissez-faire* equilibrium for a particular economy

investment choice. For instance, when $\ell = 0$ every individual gets the same utility with
an identical level of consumption, taking into account that pre-tax income equals $-\epsilon$ when
$\alpha = 1$, and zero otherwise. Since agents rank income-consumption bundles in a different
way, as a function of the investment choice, it is possible to have two solutions along the 45
degree line. Note that, despite having the same preferences over labour and consumption,
they choose different bundles of income and consumption because individual preferences
over space X also depend on the innate ability. This implies that the single-crossing property
is now only assured among those who are endowed with the same ability.

We can now formally define a social allocation $x_N \in X^n$ as incentive-compatible if and
only if

$$\text{for all } j, k \in N, \text{ either } (y_j, c_j) \left(\widehat{R}_j^{s_j}\right)^a (y_k, c_k) \text{ or } y_k > s_j^1 - \epsilon.$$

The intuition behind the first part of this constraint is rather clear. Since what matters in
the non-observable labour scenario is the amount of income that one decides to earn, every
individual chooses the available bundle that provides her with the highest level of utility.
The second part of the incentive-compatibility constraint, $y_k > s_j^1 - \epsilon$, limits the use of the
condition itself only to bundles that individual j can reach with her own skill. Note that the
allocations depicted in Fig. 4 are incentive-compatible.

As a result of this discussion about the (y, c) space, we have that any individual's well-
being must now be evaluated taking her productivity into account. Let us then present the
characterisation of the s_L^α-Implicit Transfer in the non-observable labour scenario. To do so
we focus on an example with two individuals, $j, k \in N$. Agent j has preferences $R_2 \in \mathcal{R}$
with a low aversion to labour, whereas k has preferences $R_1 \in \mathcal{R}$, such that $R_2 \succ^\ell R_1$.
Additionally, these agents are endowed with different production skills, more precisely,
$s_j^\alpha = s_H^\alpha$ and $s_k^\alpha = s_L^\alpha$. Finally, we also consider that individual k *ex post* changes to
preferences R_2 and hence regrets her initial choice. Let us assume that their initial situations
are described in the left-hand side of Fig. 5, with $T_j(z_j, R_2) > 0$ and $T_k(z_k, R_2) < 0$.
As the reference skill that characterises the comparable measure of utility is the smallest
one, individual k's well-being in space X can be assessed with the line of slope 1 (see
the right-hand side of the picture). Since individuals may transform labour into income
in different ways, the piece of notation that defines this measure states now the agent's
skill, that is $T_k^L(x_k, R_2) < 0$. In order to evaluate j's well-being this measure has to be
adjusted to take into account her higher earning ability. Specifically, her comparable pre-tax
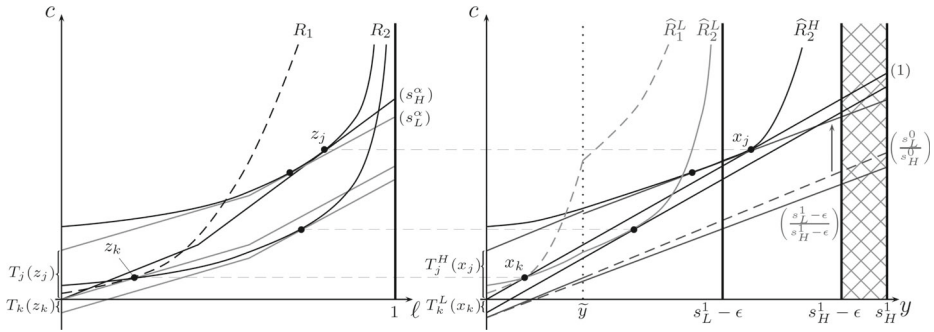income-consumption relation is defined by the ratio between the two possible consumption

Fig. 5 The s_L^α-Implicit Transfer in the second-best scenario

levels when individuals work full time. This is described by the two lines with slope $(s_L^y - \epsilon^y)/(s_H^y - \epsilon^y)$, where s^y and ϵ^y are, respectively, the skill and the investment associated with y, being y the income level in which this set of lines is tangent to the individual's indifference curve.[7]

The characterisation of any taxation policy with non-observable labour is a difficult tax because the profile of individual preferences may be excessively complex. Therefore, to be able to offer a deeper description of the second-best allocation we follow the literature on fair taxation by requiring that this profile is not overly informative (e.g., Fleurbaey and Maniquet 2011; Valletta 2014). More precisely, we consider that for any pre-tax income level lower than $s_L^1 - \epsilon$ it is not possible to identify the high-skilled agents by looking at their preferences, except those with R_F. Formally,

Assumption 1 (*No Identification*) *For every* $j \in N$ *who has preferences* $R_j \in \mathcal{R} \setminus \{R_F\}$, *there exists* $k \in N$ *endowed with* s_L^α *such that:*

$$\widehat{R}_k^L = \widehat{R}_j^{s_j}|_{[0, s_L^1 - \epsilon] \times \mathbb{R}_+}.$$

Combining this condition with the properties that we have previously endorsed implies that for low levels of income one has $\widehat{R}_f^H = \widehat{R}_{f+1}^L$, for any $R_f \neq R_F$. More importantly, under Assumption 1 the single-crossing property is recovered. Let $\widetilde{\mathcal{E}} \subset \mathcal{E}$ denote the set of economies which, besides all the previous conditions, additionally satisfies *No Identification*. Moreover, we also consider that the taxation scheme is minimal. This implies that all (y, c) points along the unique post-tax budget set are relevant, that is, each point of this set hits the indifference curve of at least one agent.

After having presented all the elements that define our framework when labour time is not observable, let us now summarise the basic outcomes that a taxation policy which aims to maximise social welfare generates.

Theorem 2 *For any economy* $e \in \widetilde{\mathcal{E}}$, *the minimal fresh start tax scheme* $\tau(y)$ *which induces a feasible and incentive-compatible allocation* $x_N \in X^n$ *that maximises the lowest* s_L^α-*Implicit Transfer across individuals is such that:*

---

[7]This measure is equivalent to the ones derived by Valletta (2014) and Fleurbaey and Valletta (2018), taking into account that they do not consider the expenditure on human capital as a part of the pre-tax income.

i)   *Perfect equality in terms of the s_L^α-Implicit Transfer cannot be achieved.*
ii) *The worst-off in terms of equivalent well-being is a low-skilled agent who makes her choice with the ex ante preferences R_1, and who either sticks to them or changes them for R_F.*
iii) *The tax function $\tau(y)$ is non-decreasing over the interval $[0, s_L^1 - \epsilon]$.*
iv) *For any feasible and incentive-compatible allocation $x'_N \in X^n$ obtained with an alternative minimal tax scheme $\tau'(y)$, we have that $x_N \mathbf{P}(e) x'_N$ whenever,*

$$\max_{y \leq s_L^1 - \epsilon} \widehat{\tau}(y) < \max_{y \leq s_L^1 - \epsilon} \widehat{\tau}'(y),$$

where $\widehat{\tau}(y)$ is the tax scheme that envelops $\widehat{R}_1^L$ and $\widehat{R}_F^L$ at bundle x_1^L.

The first conclusion of the implementation of the fresh start policy is that the full egalitarian goal is, in general, unattainable. Such an extreme result was already pointed out by Fleurbaey (2005) as a consequence of the incentive-compatible constraint. This difficulty in securing the full egalitarian goal deepens with our assumption of heterogeneity in production skills. Since the tax scheme has to treat all those who select the same income equally, any individual with a good production skill can always mimic the behaviour of a low-skilled agent, if this allows the former to benefit from an extra subsidy that is originally designed for the latter, breaking full redistribution this way. That is, the assumption of non-observable endogenous productivities exacerbates the incentive problems that one usually experiences in a second-best scenario.

The second statement of this theorem characterises the worst-off individual in terms of our comparable measure of well-being. Such a situation is determined by someone endowed with a low skill who has the *ex ante* preferences R_1, and who either sticks to her choice or experiences the highest possible regret. The fact that this minimum value may be related to a low-skilled agent who changes to R_F is rather clear. The reason why a steady individual (i.e., someone who does not regret her choice) may also be associated with the lowest individual well-being is a bit more complex. The steady low-skilled agent who shows the highest aversion to labour defines the most problematic position in terms of a potential regret. Therefore, to control future contingencies the fresh start policy leads individuals using these preferences to a situation in which they have to work more than they are willing to. Interestingly enough, these two worst-off agents may not end up with the same well-being in the final allocation. This result stems from the fact that the individual productivity is affected by the choice of labour. Consequently, if the points that define these implicit budgets are related to different investment choices, and hence different (minimal) productivities, further redistribution between the agents that characterise the minimum well-being levels may be blocked (see Fig. 7 in the appendix).

The third result in Theorem 2 describes the main role that forgiveness plays in the second-best scenario. The use of the lowest skill to define fair compensations generally yields a taxation scheme which establishes that individuals endowed with such a skill should face a zero marginal tax (see Fleurbaey and Maniquet 2011). This is so because inequalities within this segment of the population are due to different preferences over labour and consumption, something that one wants to respect when using a compensation approach. However, in the present paper we are also interested in *ex post* changes in preferences. As a result of this, and unlike previous findings, we obtain that among those with low earnings the fresh start tax scheme should be progressive. That is, apart from focusing on the agents endowed with

the lowest marginal productivity, the social preferences should also devote a share of the public resources to pay compensations to those who regret their previous choice.

The final point in Theorem 2 defines how to rank allocations in terms of social welfare. The standard literature proposes to focus, within earnings which are at most $s_L^1 - \epsilon$, on the level of income that entails the maximum tax (or the minimum subsidy). However, when introducing forgiveness we are considering that some individuals will mismanage a share of their resources modelled by the minimal tax scheme. Therefore, when comparing allocations we have to focus on the hypothetical maximum tax that a low-skilled agent would have to pay, taking into account that she may eventually waste a share of her resources. Note that this assessment of social welfare does not require an extensive knowledge of society. Only some pieces of information on the general distribution of individual characteristics are needed, specially regarding the preferences of those who enjoy the smallest level of income, i.e., x_1^L.

We can complement this characterisation of the second-best allocation with the Saez and Stantcheva's (2016) weighting incomes approach. Under this approach social preferences are represented by marginal social welfare weights that are assigned to observed earning levels. These weights capture the aggregate social gains and losses of any small tax reform. One of the interesting features of this framework is that the weights are derived from intuitions about how the final distribution of income should be according to different fairness and efficiency principles. However, Fleurbaey and Maniquet (2018) argue that this weighting incomes model, though useful, does not always allow us to characterise the optimal tax scheme. The main problem is that the weights depend on the particular allocation under consideration. Moreover, when the social ordering is relatively complex the identification of the income levels that deserve a higher weight becomes a difficult task. Finally, these authors defend that the social welfare function framework with money-metric utilities, the one we have adopted in this paper, is a flexible tool for the comparison of individual well-beings, as well as for the incorporation of other fairness considerations.

Nevertheless, since social preferences $\mathbf{R}_{s_L^\alpha}^{\text{lex}}$ involve the use of the leximin criterion many levels of earnings would have a zero weight, and hence it is relatively easy to derive the profile of social weights for our model. Computing these values allows us to minimally characterise the second-best efficient allocations under different normative principles. For instance, in a model without regret in which s_L^α is the reference skill, the social weights would be as follows (this solution is depicted in Fig. 6). First, as the full priority is put on those endowed with the lowest innate ability, weights would be set to 0 for all incomes above $s_L^1 - \epsilon$. Second, since all individuals can earn at least this quantity, in the $[0, s_L^1 - \epsilon]$ earnings range the planner should assign a strictly positive weight to that individual who pays the highest tax, and a zero weight to all other situations. Consequently, in the first part of the income distribution all individuals receive the same transfer, and hence the optimal scheme exhibits a marginal rate of taxation that is zero on incomes below $s_L^1 - \epsilon$, something represented with the thick line in Fig. 6. Above this threshold weights are null, and hence the only objective of the optimal scheme is to collect taxes to maximise social welfare, provided that the post-transfer income is non-decreasing (see the cross-hatched area in the picture). Therefore, the marginal tax in this part of the picture is non-negative.[8]

This weighting incomes model can also be used to depict the role that the fresh start ideal plays in the optimal tax scheme. In this characterisation two factors are key; namely,

---

[8]The shapes of the optimal tax schemes for alternative reference skills are described in Fleurbaey and Maniquet (2011).
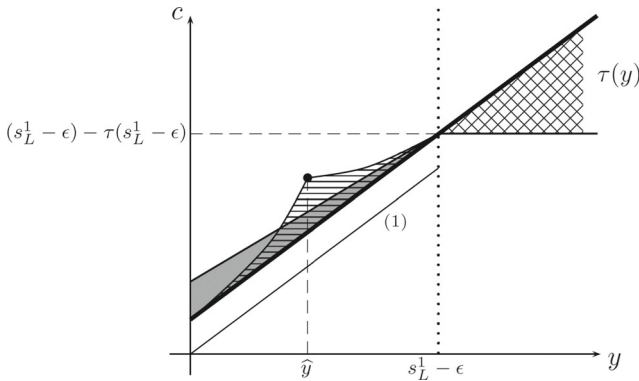
Fig. 6 Optimal second-best allocations with different weights

the fact that the highest regret is linked to preferences R_1 and the incentive-compatible constraint. On the one hand, the concern for those who may suffer the maximum regret swings the social weights towards those who are at the very end of the income distribution. On the other hand, because of incentive-compatibility the well-being of those who do not make their choices with R_1 is bounded below, and hence they will receive a zero weight. As a result of this, the shape of the optimal tax scheme would change for values lower than $s_L^1 - \epsilon$, that is, the highest income that someone endowed with the lowest skill can earn. Once again, above this threshold the only objective of the optimal scheme would be to collect taxes to maximise social welfare, provided that the post-tax budget set is non-negative. By contrast, in the $[0, s_L^1 - \epsilon]$ range the full priority is now given to those with the lowest earnings. Consequently, the budget set pivots upwards from the reference income, showing that these agents should pay a lower tax in order to be compensated for a possible change in preferences (see the shaded area in Fig. 6).

Interestingly enough, this approach can also be used to represent how the optimal tax scheme should be designed in other configurations of the forgiveness model. For example, if one considers that preferences can change in all directions, which is the only reasonable alternative to our model, the weights would be assigned in such a way that the tax schedule would induce all agents in the $[0, s_L^1 - \epsilon]$ range to get the same pre-tax income (see the lined area in Fig. 6). This is so because in this scenario there would be no reason to prioritise between the regretful who have worked too much and the regretful who have worked too little, and hence transfers related to regret would be of little use. By contrast, other configurations cannot be so easily characterised with the social weights model. This is what happens, for instance, with the libertarian approach in which differences in earning capacities are not considered unfair. In the basic configuration of this approach every individual is fully entitled to her own production, and hence the marginal tax should be equal to zero for all incomes. That is, no matter the profile of skills, a *laissez-faire* policy in which no one would be taxed should be applied, something that can be represented in Fig. 6 with the straight thin line of slope 1. However, when one includes the forgiveness ideal it is no longer clear how the shape of the optimal tax scheme should be. The planner would be facing the dilemma of letting agents be rewarded at their marginal productivity, or collecting taxes to provide a fresh start to all those who regret their previous choices.

6 Concluding remarks

The ideal of fair compensation aims to correct outcome differences that originate from unequal innate productivities, whereas the principle of forgiveness defends giving a fresh start to those who have mismanaged their initial resources. The aim of the present paper is to examine the interaction between these two ethical viewpoints. We do so in a scenario in which agents have to initially decide between investing and not investing in one's own productivity, something that may worsen the potential well-being losses that an agent suffers when regretting previous decisions. Grounded on different normative principles, we have singled out a class of social preferences that allows us to reach a compromise between the two ethical approaches. Such preferences give absolute priority to that agent with the lowest value of a particular measure of well-being that entails neither skill inequalities nor regret in the individual labour decisions.

In the second part of the paper we have evaluated the consequences of the implementation of a fresh start taxation policy that satisfies our social preferences. To do so we have defined a second-best scenario in which the planner has to design an income tax scheme under incentive-compatible constraints due to the impossibility to observe personal traits and individual labour choices. Although the analysis of any social situation in such a framework is a difficult task, we have derived a simple criterion for the comparison of different tax schemes, and hence different allocations. Such a criterion is defined in terms of the hypothetical maximum tax that would be paid among those with low incomes, assuming that they may regret their previous labour choices. We obtain that the assumption of non-observable endogenous productivities make the issue of providing individuals with the right incentives to maximise social preferences more difficult, because they have now more alternatives to replicate others' behaviour. Interestingly enough, this assessment of social welfare does not require to have an extensive knowledge of the society, but only some minimal information about those who have chosen the smallest level of income. Moreover, and unlike previous results, we find that the marginal tax rate for those who get low pre-tax earnings should optimally be non-decreasing. More precisely, provided that the tax scheme is incentive-compatible, the highest social weights should be assigned to those who earn the smallest income. This progressivity aims to respect outcome differences that are a result of elements for which agents are held responsible, and also to compensate individuals who genuinely regret being at the very low end of the income distribution.

Appendix A: Proof of Theorem 1

In order to prove this theorem we first need to introduce the following lemma:

Lemma 1 *If a social ordering function satisfies Strong Pareto, Independence and Equal Preferences Transfer, then for all $e \in \mathcal{E}$ and $z_N, z_N' \in Z^n$, if there exist $j, k \in N$ with $R_j^p = R_k^p$ such that:*

$$z_j' P_j^p z_j P_j^p z_k P_k^p z_k',$$

and $z_i = z_i'$ for all $i \neq j, k$, then $z_N \mathbf{P}(e) z_N'$.

For proof of Lemma 1 see Fleurbaey and Maniquet (2006).

Let us now turn to prove Theorem 1. The proof is split in three steps.

Step 1: For any $e \in \mathcal{E}$, let us consider two individuals $j, k \in N$ and two allocations $z_N, z'_N \in Z^n$ such that, without loss of generality, $T_k(z_k, R_k^p) = T_j(z'_j, R_j^p) > T_k(z'_k, R_k^p) = T_j(z_j, R_j^p)$, and $z_i = z'_i$ for all $i \neq j, k$. We need to prove that it must be the case that $z_N \mathbf{I}(e) z'_N$. Opposite to the desired result, let us assume that $z_N \mathbf{P}(e) z'_N$.

Let us now introduce two additional individuals b, m such that $s_b^\alpha = s_m^\alpha = s_L^\alpha$, $R_b^p = R_j^p \in \mathcal{R}$ and $R_m^p = R_k^p \in \mathcal{R}$. Let us also assume that there exist z_b, z'_b, z''_b, z'''_b, $z_m, z'_m, z''_m, z'''_m \in Z$ such that:

$$T_b(z_b, R_b^p) = T_m(z'_m, R_m^p) = T_k(z_k, R_k^p) > T_b(z'_b, R_b^p) = T_m(z_m, R_m^p) = T_j(z_j, R_j^p),$$
$$z_b = z'_j, \; z'_b = z_j, \; z_m = z'_k, \; z'_m = z_k,$$
$$z''_b I_b^p z_b, \; z'''_b I_b^p z'_b, \; z''_m I_m^p z_m, \; z'''_m I_m^p z'_m,$$
$$z''_b, z'''_b \in Z(R_b^p, s_b^\alpha), \; z''_m, z'''_m \in Z(R_m^p, s_m^\alpha),$$

where for any $R \in \mathcal{R}$ and $s^\alpha \in \mathcal{S}$, and for all $t \in \mathbb{R}$, the set $Z(R, s^\alpha) = \{z \in Z \mid z \in \max|_R B(t, s^\alpha)\}$ gathers all the bundles that maximise $R \in \mathcal{R}$ for any budget set which is defined by s^α.

According to the initial assumptions, if we apply *Separation* we can add identical individuals in both allocations without altering social preferences over them, that is, $(z_N, z_b, z_m) \mathbf{P}(e)(z'_N, z_b, z_m)$. Combining *Strong Pareto* (which implies *Pareto Indifference*) and *Equal Preferences Transfer* twice we can induce the following relations: $(z_{-\{k\}}, z'_k, z_b, z'_m) \mathbf{I}(e)(z, z_b, z_m)$ and $(z_{-\{j,k\}}, z'_j, z'_k, z'_b, z'_m) \mathbf{I}(e)(z_{-\{k\}}, z'_k, z_b, z'_m)$. If we apply *Strong Pareto* $(z'_N, z'_b, z'_m) \mathbf{I}(e)(z_{-\{j,k\}}, z'_j, z'_k, z'_b, z'_m)$. Using *Transitivity* we have that $(z'_N, z'_b, z'_m) \mathbf{P}(e)(z'_N, z_b, z_m)$, and according to *Separation* we can state that $(z'_b, z'_m) \mathbf{P}(e)(z_b, z_m)$. However, if we apply s_L^α-*Skill Transfer* and *Strong Pareto* axioms it is straightforward to obtain first $(z''_b, z''_m) \mathbf{I}(e)(z'''_b, z'''_m)$, and next that $(z_b, z_m) \mathbf{I}(e)(z'_b, z'_m)$, which yields the desired contradiction.

Step 2: For any $e \in \mathcal{E}$, let us consider, without loss of generality, two allocations $z_N, z'_N \in Z^n$ and two individuals $j, k \in N$ such that $T_j(z'_j, R_j^p) < T_k(z_k, R_k^p) < T_j(z_j, R_j^p) < T_k(z'_k, R_k^p)$, and $z_i = z'_i$ for all $i \neq j, k$. We need to prove that $z_N \mathbf{P}(e) z'_N$. Opposite to the desired result, let us assume that $z'_N \mathbf{R}(e) z_N$.

Let us now introduce two additional individuals b, m such that $s_b^\alpha = s_m^\alpha = s_L^\alpha$, $R_b^p = R_j^p \in \mathcal{R}$ and $R_m^p = R_k^p \in \mathcal{R}$. Let us also assume that there exist $z''_j, z''_k, z''_b, z''_m \in Z$, $z_b, z'_b \in Z(R_b^p, s_b^\alpha)$ and $z_m, z'_m \in Z(R_m^p, s_m^\alpha)$ such that:

$$z_j P_j^p z_b P_b^p z'_b P_b^p z''_b P_b^p z''_j P_j^p z'_j,$$
$$z'_k P_k^p z_k P_k^p z''_k P_k^p z'_m P_m^p z''_m P_m^p z_m.$$

Moreover, there exists $\delta \in \mathbb{R}_{++}$ such that:

$$c'_m - (s_m^{\alpha'_m} \ell'_m - \epsilon \alpha'_m) - \delta = c_m - (s_m^{\alpha_m} \ell_m - \epsilon \alpha_m) >$$
$$c_b - (s_b^{\alpha_b} \ell_b - \epsilon \alpha_b) = c'_b - (s_b^{\alpha'_b} \ell'_b - \epsilon \alpha'_b) + \delta.$$

According to the initial assumptions and the *Separation* axiom we have that $(z'_N, z_b, z_m) \mathbf{R}(e)(z_N, z_b, z_m)$. Applying *Lemma 1* twice we obtain that $(z'_{-\{j,k\}}, z''_j, z''_k, z''_b, z''_m) \mathbf{P}(e)(z'_N, z_b, z_m)$. Additionally, *Strong Pareto* implies that $(z'_{-\{j,k\}}, z_j, z_k, z''_b, z''_m) \mathbf{P}(e)(z'_{-\{j,k\}}, z''_j, z''_k, z''_b, z''_m)$. Finally, by *Pareto Indifference*, *Transitivity* and *Separation* we have that $(z''_b, z''_m) \mathbf{P}(e)(z_b, z_m)$. However, if we apply *Strong Pareto* we get that $(z'_b, z'_m) \mathbf{P}(e)(z''_b, z''_m)$, and according to s_L^α-*Skill Transfer* we have that $(z_b, z_m) \mathbf{R}(e)(z'_b, z'_m)$. *Transitivity* induces then that $(z_b, z_m) \mathbf{P}(e)(z''_b, z''_m)$, which yields the desired contradiction.

Finally, we can design a series of allocations that would allow us to prove that whenever there exist $z_N, z'_N \in Z^n$ such that $\min_{i \in N} T_i(z_i, R_i^P) > \min_{i \in N} T_i(z'_i, R_i^P) \Rightarrow z_N \mathbf{P}(e) z'_N$. Let us take then two allocations $z_N, z'_N \in Z^n$ such that $\min_{i \in N} T_i(z_i, R_i^P) > \min_{i \in N} T_i(z'_i, R_i^P)$.

Because of the strict monotonicity of the preferences, one can find two allocations $x_N, x'_N \in Z^n$ such that for all $i \in N$ we have $T_i(z_i, R_i^P) > T_i(x_i, R_i^P)$ and $T_i(x'_i, R_i^P) > T_i(z'_i, R_i^P)$. Moreover, there exists i_0 such that for all $i \neq i_0$:

$$T_i(x'_i, R_i^P) > T_i(x_i, R_i^P) > T_{i_0}(x_{i_0}, R_{i_0}^P) > T_{i_0}(x'_{i_0}, R_{i_0}^P).$$

Let $Q = N \setminus \{i_0\}$ and let us assume a sequence of allocations $(x_N^q)_{1 \leq q \leq |Q|+1}$ such that:

$$T_i(x_i^q, R_i^P) = T_i(x'_i, R_i^P), \quad \forall i \in Q : i \geq q,$$
$$T_i(x_i^q, R_i^P) = T_i(x_i, R_i^P), \quad \forall i \in Q : i < q,$$

whereas,

$$T_{i_0}(x_{i_0}, R_{i_0}^P) = T_{i_0}(x_{i_0}^{|Q+1|}, R_{i_0}^P) > T_{i_0}(x_{i_0}^{|Q|}, R_{i_0}^P) > \ldots > T_{i_0}(x_{i_0}^1, R_{i_0}^P) = T_{i_0}(x'_{i_0}, R_{i_0}^P).$$

This implies that $T_{i_0}(x_{i_0}^q, R_{i_0}^P) < T_{i_0}(x_{i_0}^{q+1}, R_{i_0}^P) < T_q(x_q^{q+1}, R_q^P) < T_q(x_q^q, R_q^P)$, while for all $j \neq q, i_0$, we have that $T_j(x_j^q, R_j^P) = T_j(x_j^{q+1}, R_j^P)$. As we have previously proved, it must be the case that $x_N^{q+1} \mathbf{P}(e) x_N^q, \forall q \in Q$. According to the initial assumptions, by *Strong Pareto* we have that $z_N \mathbf{P}(e) x_N^{|Q|+1}$ and $x_N^1 \mathbf{P}(e) z'_N$. Finally, by *Transitivity* we have that $z_N \mathbf{P}(e) z'_N$.

Step 3: In this last step of the proof we extend our result in order to meet the leximin criterion, that is, for any $e \in \mathcal{E}$ and $z_N, z'_N \in Z^n$ we have to show that $(T_i(z_i, R_i^P))_{i \in N} >_{\text{lex}} (T_i(z'_i, R_i^P))_{i \in N} \Rightarrow z_N \mathbf{P}(e) z'_N$. Without loss of generality, let us assume that there exist $j, k \in N$ such that $\min_{i \in N} T_i(z_i, R_i^P) = T_j(z_j, R_j^P) = T_k(z'_k, R_k^P) = \min_{i \in N} T_i(z'_i, R_i^P)$. Additionally, we consider that $\min_{i \in N} T_i(x_i, R_i^P) > \min_{i \in N} T_i(x'_i, R_i^P)$, where $x_N, x'_N \in Z^{n-1}$ are the allocations that are constructed as the original ones but removing, in each one of them, the individual with the lowest s^α-Implicit Transfer. By *Strong Pareto* and repeated application of Step 1 we have that $\overline{z}_N \mathbf{I}(e) z_N$, where $\overline{z}_N$ is designed such that the lexicographic values of the implicit transfers in both allocations is the same, that is $(T_i(\overline{z}_i, R_i^P))_{i \in N} =_{\text{lex}} (T_i(z_i, R_i^P))_{i \in N}$, and moreover $\min_{i \in N} T_i(\overline{z}_i, R_i^P) = T_k(\overline{z}_k, R_k^P) = T_k(z'_k, R_k^P)$. Then, by *Separation* and Step 2 we have that $\overline{z}_N \mathbf{P}(e) z'_N$, and by means of *Transitivity* we reach the desired result $z_N \mathbf{P}(e) z'_N$.

Appendix B: Proof of Theorem 2

This proof is based on the result obtained by Fleurbaey (2005). Let us start the proof by showing that, for any economy $e \in \widetilde{\mathcal{E}}$, in any *ex post* optimal incentive-compatible allocation $x_N = (y_i, c_i)_{i \in N} \in X^n$ all resources are exhausted, that is, $\sum_{i \in N} \tau(y_i) = 0$. Opposite to the desired result, let us consider an incentive-compatible allocation $x'_N = (y'_i, c'_i)_{i \in N} \in X^n$ with $\sum_{i \in N} \tau(y'_i) > 0$.

If $x'_j = x'_k$ for all $j, k \in N$, it is possible to define a parameter $\delta > 0$ such that if we replace the original allocation x'_N by $x_N^\delta = (y'_i, c'_i + \delta)_{i \in N} \in X^n$, we obtain a new feasible

and incentive-compatible allocation in which, because of strict monotonicity, all individuals are better-off.

Let us now deal with the case in which individuals do not have the same bundle, that is, there exist $j, k \in N$ such that $x'_j \neq x'_k$. Since the feasible space is defined in terms of pre-tax earnings and final consumption, because of the incentive-compatible constraint the final allocation must be distributed along the unique (net of tax) budget set. Note that due to the monotonicity of the preferences and the incentive-compatible constraint we have to restrict our attention to tax functions such that $y - \tau(y)$ is non-decreasing. Hence, those who choose the largest level of income also have the highest consumption. Moreover, due to the assumptions of the model and *No Identification*, among these agents we can find one endowed with both preferences R_F and the highest innate ability.

Therefore, because of the single-crossing property we can use the extra resources to slightly improve the situation of this agent, increasing, this way, social welfare due to the lexicographic condition of the social preferences. Let $x^s_f \in X$ denote the choice of an agent who has a production skill $s^\alpha \in \{s^\alpha_L, s^\alpha_H\}$ and the *ex ante* preferences $R_f \in \mathcal{R}$. Then, by increasing the consumption associated with bundle x^H_F social welfare increases. If there exist incentive-compatibility problems with other agents who are indifferent to x^H_F, we can always move the bundle along the indifference curve of these individuals, if they share preferences. Note that this movement does not provide the other agents with incentives to change their own choice. As a result of this, in this new feasible allocation incentive-compatibility is still satisfied and social welfare is higher.

Let us now proceed to prove the four different results of Theorem 2.

i)   For any economy $e \in \widetilde{\mathcal{E}}$, if all individuals are equal-off in terms of the reference comparable well-being measure, then any $j, k \in N$ who share *ex post* preferences $R_f \in \mathcal{R}$ will have their bundles located in the same indifference curve in space Z. Consequently, in the income-consumption space X their indifference curves have to start from the same point in the vertical axis (zero income point). If these agents are endowed with different production skills, from this point onwards the curve associated with $\widehat{R}^H_f$ has to be located to the right of that of $\widehat{R}^L_f$. This is so because the high-skilled individual can always get the same (y, c) pair as the low-skilled one by using a lower labour time.

As a result of this, and because of the single-crossing property, to satisfy the incentive-compatible constraints all bundles but x^H_F must be identical and associated with $y = 0$. Additionally, this zero income point and x^H_F have to be located along the same $\widehat{R}^H_F$ indifference curve. Therefore, due to the single-crossing property, the s^α_L-Implicit Transfers can never be identical unless all individuals show an extremely low concern for labour.

ii) Let us start the proof by considering the group of steady agents (i.e., those who do not regret their initial choice) who are endowed with a good innate ability. By incentive-compatibility they can never be worse-off than the steady bad production skill individuals who have the same concern for labour. Let us now analyse this second group of agents. Due to the incentive-compatible constraint and the single-crossing property, those with $\widehat{R}^L_f \neq \widehat{R}^L_1$ are better-off, or at least equal-off, than the regretful low-skilled agents who choose with preferences $\widehat{R}^L_{f_1}$, where $\widehat{R}^L_f \succ^\ell \widehat{R}^L_{f_1}$, and who *ex post* substitute them for $\widehat{R}^L_f$.

Consequently, let us now focus on the set of individuals who regret their initial choice. As regards the high-skilled ones, for any $R_f, R_{f_1} \in \mathcal{R}$, where $R_f \succ^\ell R_{f_1}$,
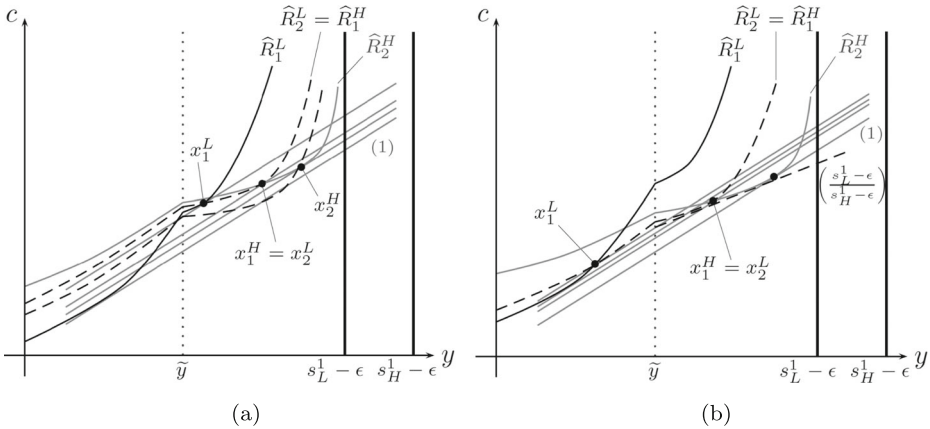
Fig. 7 Proof of Theorem 2.*ii*

there exists a third type $R_{f_2} \in \mathcal{R}$ with $R_{f_2} \succ^\ell R_f$ such that, because of the incentive-compatible constraint and the single-crossing property, $T_{f_2}^H(x_f^H) \geq T_{f_2}^H(x_{f_1}^H)$.[9] Note that this argument is valid for all agents but those making their choice with R_1. This same line of reasoning can be applied to the bad production skill agents.

Additionally, note that by Assumption 1 we have that $y_1^H \geq y_1^L$ and $c_1^H \geq c_1^L$, where y_f^s (respectively, c_f^s) denotes the income choice (respectively, consumption choice) of an agent who has production skill $s^\alpha \in \{s_L^\alpha, s_H^\alpha\}$ and *ex ante* preferences $R_f \in \mathcal{R}$. Therefore, it has to be the case that $T_F^H(x_1^H) \geq T_F^L(x_1^L)$.

In consequence, in order to characterise the lowest s_L^α-Implicit Transfer we have to focus on the low-skilled individuals with the *ex ante* preferences $R_1 \in \mathcal{R}$. However, because of the single-crossing property, for any $R_f \in \mathcal{R}$ we have either $T_F^L(x_1^L) \leq T_f^L(x_1^L)$ or $T_1^L(x_1^L) \leq T_f^L(x_1^L)$.

Finally, Fig. 7 depicts two examples in which either $T_1^L(x_1^L) < T_F^L(x_1^L)$ or $T_1^L(x_1^L) > T_F^L(x_1^L)$. In both cases all incentive-compatibility constraints are satisfied in such a way that it is not possible to increase social welfare.

iii) In order to prove the third point of Theorem 2, let us first show that for every $R_f \in \mathcal{R} \setminus \{R_F\}$ one has that $x_{f+1}^L \widehat{I}_{f+1}^L x_f^L$.

Opposite to the desired result, let us assume that $x_{f+1}^L \widehat{P}_{f+1}^L x_f^L$ for some $R_f \neq R_F$. Then, due to the properties of the model we have that $x_{f_1}^L \widehat{P}_{f_1}^L x_{f_0}^L$ for all $f_1 > f \geq f_0$.

Let us now consider that for every $f_2 > f$ there exists $f_3 > f$ such that $x_2^L \neq x_3^L$ and $x_2^L \widehat{I}_2^L x_3^L$. If $f_2 < f_3$, by incentive-compatibility we have $x_3^L \neq x_4^L$ and $x_3^L \widehat{I}_3^L x_4^L$ with $f_3 < f_4$, which yields an impossibility since, eventually, it is not possible to obtain the same relation for preferences R_F. Note that we may have $x_F^L \widehat{I}_F^L x_F^H$, with $x_F^L \neq x_F^H$. However, in that case all bundles would be in the $[0, s_L^1 - \epsilon]$ interval, and hence the previous argumentation would lead us to the same difficulty, but including

[9] $T_f^s(x_{f'}^{\bar{s}})$ represents the well-being level that an agent endowed with $s^\alpha \in \{s_L^\alpha, s_H^\alpha\}$ and *ex post* preferences $R_f \in \mathcal{R}$ gets from a bundle $x_{f'}^{s'} \in X$, which was chosen by someone who has a production skill $\bar{s}^\alpha \in \{s_L^\alpha, s_H^\alpha\}$ and *ex ante* preferences $R_{f'} \in \mathcal{R}$.

all those endowed with a good production skill. If $f_2 > f_3$ a similar impossibility occurs since we are constrained by $x_{f_1}^L \widehat{P}_{f_1}^L x_f^L$.

Therefore, under the previous assumption there exits $f_2 > f$ such that for all $f' \neq f_2$ we have that $x_{f_2}^L \widehat{P}_{f_2}^L x_{f'}^L$, with $x_{f'}^L \neq x_{f_2}^L$. Hence, we can slightly increase the tax associated with this bundle, and any other one that is identical, leaving the other choices unchanged. This yields a new incentive-compatible allocation in which some resources are saved without altering the lowest s_L^α-Implicit Transfer, which has to be related to a bundle chosen with R_1 (see point ii). These resources can be used to increase the minimum equivalent well-being (see Fleurbaey 2005). Note that if both $x_{f_2}^L = x_{f_2-1}^H$ and $x_{f_2-1}^H \widehat{I}_{f_2-1}^H x_{f_2}^H$ with $y_{f_2}^H > s_L^1 - \epsilon$, this may prevent us from obtaining the desired result. In such a case we can also increase the tax paid in $x_{f_2}^H$, unless we also have that $x_{f_2}^H \widehat{I}_{f_2}^H x_{f_2+1}^H$. However, this would imply that $x_{f'}^H \widehat{P}_{f'}^H x_{f''}^H$, with $y_{f'}^H > s_L^1 - \epsilon$, for all $f' > f_2 \geq f''$, which, as we have previously argued, leads to an impossibility.

As a result of this, for any $R_f \in \mathcal{R} \setminus \{R_F\}$ it has to be the case that $x_{f+1}^L \widehat{I}_{f+1}^L x_f^L$.

Let us finally consider that for some $f < F$ one has $\tau(y_f^L) > \tau(y_{f+1}^L)$, where $y_f^L < y_{f+1}^L$. Then, since $x_{f+1}^L \widehat{I}_{f+1}^L x_f^L$, if we change the bundle x_{f+1}^L by x_f^L we obtain a new allocation that both is incentive-compatible and saves resources. Moreover, the lowest s_L^α-Implicit Transfer does not decrease. Consequently, in the optimal allocation one has $\tau(y_f^L) \leq \tau(y_{f+1}^L)$ for all $f < F$.

$iv)$ Let us finish the proof of Theorem 2 by describing how to compare any two different minimal taxation schemes in terms of social welfare.

For any $i \in N$, the s_L^α-Implicit Transfer in the first-best scenario is defined as:

$$T_i(z_i, R_i^p) = \max\{t \in \mathbb{R} \mid \forall(\ell, c) \in Z \text{ s.t. } c \leq \max\{s_L^0 \ell, s_L^1 \ell - \epsilon\} + t, \ z_i R_i^p(\ell, c)\},$$

which is equivalent to write that:

$$T_i(z_i, R_i^p) = \min\{t \in \mathbb{R} \mid \exists(\ell, c) \in Z \text{ s.t. } c = \max\{s_L^0 \ell, s_L^1 \ell - \epsilon\} + t, \ z_i I_i^p(\ell, c)\}.$$

Since we are analysing the fresh start taxation policy in a second-best scenario, such a measure has to be converted to the non-observable labour space X. Specifically:

$$T_i^{s_i}(x_i, R_i^p) = \min\{t \in \mathbb{R} \mid \exists(y, c) \in X \text{ s.t. } c = \left(\frac{s_L^y - \epsilon^y}{s_i^y - \epsilon^y}\right) y + t, \ x_i (\widehat{I}_i^{s_i})^p (y, c)\}$$
$$= \min\left\{c - \left(\frac{s_L^y - \epsilon^y}{s_i^y - \epsilon^y}\right) y \mid (y, c) \in X, \ x_i (\widehat{I}_i^{s_i})^p (y, c)\right\},$$

which for a low-skilled agent turns into the following expression:

$$T_i^L(x_i, R_i^p) = \min_{(y,c)(\widehat{I}_i^L)^p x_i} (c - y).$$

By the second point of Theorem 2 we know that the worst-off individual is a low-skilled agent who either sticks to R_1 or changes such preferences for R_F, and hence,

$$\min_{i \in N} T_i^{s_i}(x_i, R_i^p) = \min_{(y,c) \in \widehat{\tau}(y), \ y \leq s_L^1 - \epsilon} (c - y),$$

where $\widehat{\tau}(y)$ is the set of (y, c)-points that envelops the indifference curves of $\widehat{R}_1^L$ and $\widehat{R}_F^L$ which pass through x_1^L, and that are induced by the unique budget set modified by the minimal tax scheme $\tau(y)$. Therefore:

$$\min_{i \in N} T_i^{s_i}(x_i, R_i^p) = \min_{y \le s_L^1 - \epsilon} -\widehat{\tau}(y).$$

Finally, for any pair $x_N, x'_N \in X^n$ obtained, respectively, with minimal tax schemes $\tau(y)$ and $\tau'(y)$, we know that $\min_i T_i^{s_i}(x_i, R_i^p) > \min_i T_i^{s_i}(x'_i, R_i^p)$ implies that allocation x_N is socially preferred to x'_N. Therefore, by applying our previous finding and basic algebra we have that:

$$\max_{y \le s_L^1 - \epsilon} \widehat{\tau}(y) < \max_{y \le s_L^1 - \epsilon} \widehat{\tau}'(y) \Rightarrow x_N \mathbf{P}(e) x'_N.$$

References

Arneson, R.J.: Equality and equal opportunity for welfare. Philos. Stud. **56**(1), 77–93 (1989)

Arrow, K.J.: Social Choice an Individual Values. Wiley, New York (1951)

Bossert, W., Fleurbaey, M., Van de gaer, D.: Responsibility, talent, and compensation: A second-best analysis. Rev. Econ. Des. **4**(1), 35–55 (1999)

Bovenberg, A.L., Jacobs, B.: Redistribution and education subsidies are siamese twins. J. Public Econ. **89**(11), 2005–2035 (2005)

Calo-Blanco, A.: Fairness, freedom, and forgiveness in health care. Soc. Choice Welf. **43**(1), 141–151 (2014)

Calo-Blanco, A.: Fair compensation with different social concerns for forgiveness. Rev. Econ. Des. **20**(1), 39–56 (2016)

Calo-Blanco, A.: Health, responsibility and taxation with a fresh start. SERIEs **8**(2), 177–199 (2017)

Cohen, G.A.: On the currency of egalitarian justice. Ethics **99**(4), 906–944 (1989)

D'Aspremont, C., Gevers, L.: Equity and the informational basis of collective choice. Rev. Econ. Stud. **44**(2), 199–209 (1977)

Dworkin, R.: What is equality? part 1: Equality of welfare. Philos. Public Aff. **10**(3), 185–246 (1981a)

Dworkin, R.: What is equality? part 2: Equality of resources. Philos. Public Aff. **10**(4), 283–345 (1981b)

Dworkin, R.: Sovereign Virtue: The Theory and Practice of Equality. Stanford series in philosophy. Harvard University Press, Cambridge (2000)

Dworkin, R.: Sovereign virtue revisited. Ethics **113**(1), 106–143 (2002)

Findeisen, S., Sachs, D.: Education and optimal dynamic taxation: The role of income-contingent student loans. J. Public Econ. **138**, 1–21 (2016)

Fleurbaey, M.: Freedom with forgiveness. Polit/ Philos. Econ. **4**(1), 29–67 (2005)

Fleurbaey, M.: Fairness, Responsibility, and Welfare. Oxford University Press, Oxford (2008)

Fleurbaey, M., Maniquet, F.: Fair social orderings when agents have unequal production skills. Soc. Choice Welf. **24**(1), 93–127 (2005)

Fleurbaey, M., Maniquet, F.: Fair income tax. Rev. Econ. Stud. **73**(1), 55–83 (2006)

Fleurbaey, M., Maniquet, F.: A Theory of Fairness and Social Welfare. Cambridge University Press, Cambridge (2011)

Fleurbaey, M., Maniquet, F.: Optimal income taxation theory and principles of fairness. J. Econ. Lit. **56**(3), 1029–1079 (2018)

Fleurbaey, M., Valletta, G.: Fair optimal tax with endogenous productivities. J. Public Econ. Theory **20**(6), 849–873 (2018)

Hansson, B.: The independence condition in the theory of social choice. Theor. Decis. **4**(1), 25–49 (1973)

Jacobs, B., Bovenberg, A.L.: Optimal taxation of human capital and the earnings function. J. Public Econ. Theory **13**(6), 957–971 (2011)

Luttens, R.I., Ooghe, E.: Is it fair to 'make work pay'? Economica **74**(296), 599–626 (2007)

Mirrlees, J.A.: An exploration in the theory of optimum income taxation. Rev. Econ. Stud. **38**(2), 175–208 (1971)

Pazner, E.A.: Equity, nonfeasible alternatives and social choice: a reconsideration of the concept of social welfare. In: Laffont, J.J. (ed.) Aggregation and Revelation of Preferences. North-Holland, Amsterdam (1979)

Rawls, J.: A Theory of Justice. Harvard University Press, Cambridge (1971)

Roemer, J.E.: Equality of Opportunity. Harvard University Press, Cambridge (1998)

Saez, E., Stantcheva, S.: Generalized social marginal welfare weights for optimal tax theory. Am. Econ. Rev. **106**(1), 24–45 (2016)

Schokkaert, E., de gaer, D.V., Vandenbroucke, F., Luttens, R.I.: Responsibility sensitive egalitarianism and optimal linear income taxation. Math. Soc. Sci. **48**(2), 151–182 (2004)

Stantcheva, S.: Optimal taxation and human capital policies over the life cycle. J. Polit. Econ. **125**(6), 1931–1990 (2017)

Valletta, G.: Health, fairness and taxation. Soc. Choice Welf. **43**(1), 101–140 (2014)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.