Economia Politica
https://doi.org/10.1007/s40888-022-00290-8

ORIGINAL PAPER

Optimal redistribution behind the veil of ignorance

Antonio Abatemarco[1] · Francesca Stroffolini[2]

Received: 22 July 2021 / Accepted: 12 December 2022

Abstract
We propose a formalization of the Difference Principle (maximin) by which Rawls' contribution is shown to go beyond distributive value judgments in such a way as to embrace efficiency issues as well. In our model, inequalities are shown to be permitted as far as they stimulate a greater effort in education (or training), and so economic growth. This is the only possibility for an income disparity to be unanimously accepted by both the most-, and above all, the least-advantaged individual. In this vein, we highlight the peculiarity of the Rawlsian equity-efficiency trade off behind the veil of ignorance. Finally, by recalling the old tradition of 'universal ex-post efficiency', we identify the set of Rawls-optimal social contracts, which is shown to be a subset of Pareto-optimal ones.

Keywords Social contract · Equity · Efficiency · Rawls

JEL Classification D63 · D31 · J31

1 Introduction

According to Rawls' Theory, for any socioeconomic inequality to be regarded as 'legitimate', this disparity must be beneficial to the least-advantaged individual in such a way as to be unanimously accepted by both the least- and the most-advantaged

We are grateful to Alberto Bennardo, Gaetano Bloise, Marc Fleurbaey, Giuseppe Russo, Mario Tirelli for very helpful comments and suggestions. The usual disclaimers apply.

✉ Antonio Abatemarco
aabatemarco@unisa.it

Francesca Stroffolini
stroffol@unina.it

[1] Department of Economics and Statistics, CELPE, University of Salerno, Via Giovanni Paolo II, 132-84084 Fisciano, SA, Italy

[2] Department of Economics and Statistics, University of Napoli "Federico II", Complesso Universitario di Monte Sant'Angelo, Via Cintia, 21-80126 Naples, Italy

Published online: 06 January 2023

individual. This is known as Difference Principle (maximin) and is clearly essential for social cohesion.

The Difference Principle has been transposed in economic theory by claiming that, according to Rawlsian social welfare function, an allocation is to be preferred if and only if the 'least-advantaged' individual is better off; this is the main idea usually ascribed to the *maximin* principle as represented by the well known Leontief preferences (Alexander 1974). However, to the extent that efficiency issues are totally disregarded behind the veil of ignorance, any attempt to give an economic interpretation to Rawls' Theory of Justice (hereafter, Theory) would inevitably fail (e.g. Roemer and Trannoy (2016))[1]; most importantly, as far as economic incentives are disregarded, it wouldn't be clear why the least-advantaged individual should be willing to be penalized by unequal sharing of the cake.

In this paper, we propose a formalization of the Difference Principle where the identification of legitimate inequalities depends on the impact of inequality on individual incentives to effort in education. Intuitively, in our model distributive aspects are assumed to impact on growth to the extent that the magnitude of income inequalities is expected to influence individual incentives to effort in education, and so productivity in the labor market. In this context, it may be well the case that the least-advantaged individual accepts the income disparity if s/he is more than compensated by economic growth originating from better incentives to effort for the entire community.

In our view, this is a better starting-gate to import Rawls contribution in economic theory, since the Difference Principle goes well beyond the proposal of a distributive value judgment in such a way as to embrace efficiency issues as well. This seems to be consistent with Rawls' reply to Musgrave (1974) clarifying that "*it is not correct. that maximin gives no weight to efficiency. It imposes a rule of functional contribution among inequalities; and since it applies to social arrangements that are mutually advantageous, some weight is given to efficiency*" (Rawls 1974, p. 648).

In our model, we analyze the possibility of two parties (souls) to agree on a social contract behind the veil of ignorance, whose economically relevant output consists of the "scheme of wages" intended as the (linear) redistributive function associating a wage rate (e.g., hourly wage) to each productivity level (hereafter, earnings capacity) in the working life. With this purpose in mind, we consider a three-stages sequential game, where information on (i) ambitions (preference type) and (ii) native talent (abilities) is progressively acquired over time.

The timing of the model is defined as follows. At time 0 (original position), when no information is available on either (i) preference type, or (ii) native talent, individuals—actually, *souls* at this stage—agree unanimously on the social contract implying some redistribution of individual earnings capacity. At time 1 (educational stage), individual preferences are revealed and the effort decision (in education) is taken under uncertainty conditions on native talent. Most importantly, in line with Rawlsian spirit,

[1] Not surprisingly, in the *Preface* of the Restatement published 30 years after the Theory (Rawls 2001, p. xv), Rawls claims: "*In this work I... rectify the more serious faults in A Theory of Justice that have obscured the main ideas of justice as fairness, as I called the conception of justice presented in that book.*".

the native talent is assumed to become (indirectly) observable at the working stage only, since it is not measurable ex-ante and strongly influenced by the shape of social institutions revealing ex-post only. At time 2 (working stage), the realized talent, i.e. the earnings capacity (or productivity) achieved as a result of native talent endowment and effort in education, is observable and the scheme of wages (social contract) agreed behind the veil of ignorance applies.

Solving by backward induction, we show that multiple optimal social contracts may exist behind the veil of ignorance since the individual with the higher propensity to effort in education (preference type) might be associated, ex-post, either to the better, or to the worse endowment in terms of native talent.

In order to manage uncertainty conditions concerning the matching between individual preferences and talent types, given our formalization of the Difference Principle, we define the *set of optimal contracts* behind the veil of ignorance according to the notion of ‘universal ex-post efficiency’ (Starr 1973; Harris 1978; Hammond 1981). In this vein, an allocation is said to be *universally ex-post Rawls-optimal* if there is no other allocation by which the earnings capacity of the least-advantaged individual can be improved in *each* state. Within this framework, we confirm the standard result by which Rawls-optimal contracts must be a subset of Pareto-optimal ones. In addition, we highlight that social contracts are more redistributive at the optimum when individuals differ more in terms of earnings capacity.

Our contribution is twofold. First, we prove that a time-consistent formalization of the Difference Principle is feasible, provided that information on preference type and native talent is assumed to reveal progressively in a three-stage sequential setting, with native talent becoming observable at the last stage only. Second, we show that the set of Pareto-optimal contracts (under uncertainty conditions) can be refined according to Rawlsian distributive justice, while preserving unanimity conditions behind the veil of ignorance.

The paper is organized as follows. In Sect. 2, we present the major traits of the Theory by recalling the definition of the original position and the two Rawlsian principles, respectively, the Liberty and the Equality principle. The basic framework of our model, as well as the optimal decision of effort in education, is discussed in Sect. 3. In Sect. 4, the set of optimal social contracts is derived under uncertainty conditions. The major novelties of our model, as compared to common understanding of Rawls’ Theory in economics, are discussed in Sect. 5.

2 The theory of justice

Rawls’ (1971) Theory is usually accommodated in the old tradition of *social contractualism* whose best known proponents are Locke and Rousseau. Specifically, Rawls explores the possibility of a social contract, with specific distributive value judgments, to be agreed in the original position (behind the veil of ignorance), when *“individuals view themselves as potential occupants of each position in society”* (Saposnik 1981). The possibility of an unanimous agreement on a social contract is investigated ex-ante—in a sort of Constitutional stage—so as to bypass the need for unanimity conditions ex-post; quoting (Rousseau 1762, p.6), “*[t]he law of majority* *voting is itself something established by convention, and presupposes unanimity, on one occasion at least*".

In what follows, we discuss the major traits of the Theory of Justice which are indispensable for the comprehension of the model presented in Sect. 3. Thus, we briefly characterize the Rawlsian original position and the two principles of justice.

2.1 Contractualism in the original position

The focus on the original position is key in the Theory as, it is said, in order to permit a *fair* agreement (hence, the name Justice as Fairness) between free and equal persons, contractualism is required to abstract from contingencies—the particular features and circumstances of persons—which would inevitably introduce bargaining advantages jeopardizing the possibility of an overlapping consensus, and so the stability of the political institutions.[2]

Three conditions—fundamental for our formalization of the Difference Principle—are said to characterize the original position, "*(a) the parties do not have any knowledge of their desires and ends (except what is contained in the thin theory of the good, which supports the account of primary goods)...; (b) they do not know, and a fortiori cannot enumerate, the social circumstances in which they may find themselves or the array of techniques their society may have at its disposal; and (c) even if they could enumerate these possibilities, they have no grounds for relying on one probability distribution over them rather than another...*" (Rawls 1974, p.649).

First (a), individuals may differ from each other in terms of preference type—or 'ambitions' in Rawls' words—but these are unknown behind the veil of ignorance. Second (b), individuals may also differ with respect to both social circumstances (e.g., social class of origin) and natural circumstances (e.g., native talent) but, once again, this information has not revealed yet in the original position. Most importantly, to the extent that 'techniques at disposal of the society' are unknown at this stage, native talent is said to be merely potential and not measurable apart from social institutions revealing ex-post; e.g., the same native endowment (say intelligence, greed, artistic talent, etc...) may be more or less successful in the society depending on social and other contingencies.[3] Third (c), the social contract is agreed under uncertainty conditions when the lack of information is so radical that probabilities can be only defined in classical terms; i.e., since nothing makes one case more frequent than any other, each case is to be considered as equally possible.

Altogether, by excluding information (a-b-c), it must be the case that, in the original position, no one is advantaged or disadvantaged by natural chances or social contingencies. This is a *conditio sine qua non* (impartiality) for the social contract

[2] As far as 'freedom from personal interests and desires' is said to be a *conditio sine qua non* for any definition of justice to be valuable, Rawls is usually recognized as Kantian (Hampton 1980).

[3] "[T]he conceptions of the good that individuals form depend in part on their natural abilities and the way in which these are shaped and realized by social and other contingencies" (Rawls 1975, p.552). For an extensive discussion on the non-measurability of native talent behind the veil of ignorance, see Rawls' (1974) reply to Alexander and Musgrave.

to achieve an overlapping consensus across moral persons and generations, so as to grant the stability of political institutions.

What is known behind the veil of ignorance, instead, is the set of 'valuables' to be considered in the social contract, i.e. *primary goods*. Indeed, individuals are assumed to agree on the identification of primary goods which, according to Rawls, consist of those things citizens need, as free and equal persons, in order to have 'command' over exchangeable means for satisfying human needs and interests, and which have not to be confused with things it is simply rational to want or desire, or to prefer or even to crave. In this sense, individual preferences are assumed to be concerned with the instrumental value of goods—i.e. 'command over resources' according to Sugden (1993)—more than their intrinsic value.[4] In this perspective, for instance, income and wealth are said to belong to the set of primary goods to the extent that they make a person capable of pursuing his/her own interests and of being a fully cooperating member of the society.[5]

Within the Restatement (Rawls 2001), most of the emphasis is posed on the lifetime earnings capacity—or, 'lifetime income prospect' in Rawls' words—which is intended as a synthetic measure, or index, quantifying the primary goods an individual may have access to when the working age is achieved. Most importantly, the lifetime earnings capacity is a potential value which is defined up to the entire time endowment in the labor market, leisure included, of each individual.[6] This is crucial in Rawls' thought. Provided that primary goods available to each individual incorporate an equal time endowment for everyone, it must be the case that the sole income inequalities due to different earnings capacity (i.e. productivity) matter, and not those originating from different effort decisions in the labor market; in other words, income inequalities among individuals with the same earnings capacity, but different effort decisions in the labor market (to be not confused with effort in education) are taken as legitimate in Rawls' perspective.

According to Rawls, differences in citizens' earnings capacity are influenced by (i) their native endowments, (ii) their preferences, (iii) their opportunities for education, and (iv) their good or ill fortune over the course of life. In this perspective, if

[4] Notably, Rawls' focus on primary goods — intended as instrumental for opportunities—has been strongly criticized by Sen (1992). According to the latter, it is more appropriate to value opportunities directly—as the capability approach does—rather than focusing on primary goods which have no intrinsic value independently from the opportunities they give.

[5] "I note some possible misinterpretations of primary goods that may lead one to overemphasize their individualistic bias. First: a comment about wealth.. wealth consists of (legal) command over exchangeable means for satisfying human needs and interests... For whatever form they take, natural resources and the means of production, and the rights to control them, as well as rights to services, are wealth" Rawls (1975, p.540).

[6] "In elaborating justice as fairness we assume that all citizens are normal and fully cooperating members of society.. [and so] willing to work and to do their part in sharing the burdens of social life, provided of course the terms of cooperation are seen as fair. But how is this assumption expressed in the Difference Principle?... Are the least advantaged, then, those who live on welfare and surf all day off Malibu? This question can be handled in two ways: one is to assume that everyone works a standard working day; the other is to include in the index of primary goods a certain amount of leisure time... Surfers must somehow support themselves. Of course, if leisure time is included in the index, society must make sure that opportunities for fruitful work are generally available" (Rawls 2001, p.179).

education opportunities are universally granted independently from the social class of origin (which is known as Rawlsian Fair Equality of Opportunity), then the wage rate paid in the labor market is expected to reflect the earnings capacity achieved as a result of the endowment in terms of native talent and the effort investment in education.

In this framework, Rawls proposes the application of a redistributive system to be agreed behind the veil of ignorance, namely the *scheme of wages*, which redistributes the earnings capacity achieved by individuals. In this way, the wage rate paid to each individual in the labor market does no longer need to coincide with his/her earnings capacity, but it is defined as a combination of his/her own earnings capacity and the one realized by the others, in a way that embodies some redistribution from the most to the least-advantaged (as identified in terms of earnings capacity). Hence, in the Rawlsian *well-ordered* society, all of the citizens contribute "*to the good of others by training and educating their native endowments and putting them to work within a fair system of cooperation*" (Rawls 2001, p.68).

2.2 The two principles of justice

Rawls' theory is primarily concerned with the political consensus on basic principles which are implemented to order a society in such a way as to permit the sole '*just*' inequalities ('*not unjust*' in Rawls' words).[7] Given the very basic set up characterizing the original position, Rawls suggests two principles which, in his view, would make disparities in earnings capacity just, i.e. legitimate and consistent with the idea of free and equal citizenship in a society seen as a fair system of cooperation: the principle of Liberty and the principle of Equality.

According to the Liberty principle, "[e]*ach person has an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for all*" (Rawls 1974, p.639). By the principle of Equality, "[s]*ocial and economic inequalities are to meet two conditions: they must be (a) to the greatest expected benefit of the least-advantaged (the maximin criterion); and (b) attached to offices and positions open to all under conditions of fair equality of opportunity*" (Rawls 1974, p.639).

The Liberty principle is said to have a priority on Equality, meaning that the former cannot be violated in the name of the latter. Such a priority is crucial for any attempt to formalize Rawls' thought. For instance, it automatically implies that equality cannot be pursued through progressive income taxation as this would violate the Liberty principle; the redistribution of wealth and income through progressive taxation—which is different from the redistribution operated through the scheme of wages agreed behind the veil of ignorance—can be admitted exclusively to prevent excessive concentrations of property and wealth, especially those likely

[7] In terms of opportunity egalitarianism, this implies that the emphasis is posed on principles, and not on the metaphysics of the *equalizandum* which strongly characterizes Sen's ideal of equality of opportunity (Sugden 1993).

to lead to political domination, as they would threaten the political liberties, i.e. the basic liberties safeguarded by the first principle.[8]

The second principle—the Equality principle—embodies two different criteria, respectively, (a) the 'Difference Principle' and (b) the principle of 'Fair Equality of Opportunity', where the latter is said to have a priority on the former. Since we are concerned with the formalization of the sole Difference Principle, in our model we assume that Fair Equality of Opportunity holds true, implying that educational opportunities are already granted by social institutions to all members of the society. In this sense, individuals are assumed to play in open and competitive markets, where access to offices and positions is universally granted. Hence, the social contract is to be designed in such a way as to respect the Difference Principle, that is, to permit the sole inequalities of wage rates benefitting the least-advantaged individual. Remarkably, a criterion is defined behind the veil of ignorance by which the least-advantaged is 'identified' ex-post only, that is, once the earnings capacity has revealed.

To the extent that political institutions are supposed to neutralize different opportunities for education (and, given that good or ill fortune is normally distributed), it must be the case that, under Fair Equality of Opportunity, citizen's disparities of earnings capacity may originate exclusively from different native talent and/or preference type (ambitions).

Most importantly—and to our knowledge this aspect has not been properly emphasized in the common understanding of Rawls' thought in economics—worse endowments in terms of native talent (whose identification is possible only ex-post when the shape of social institutions has revealed) do not necessarily imply lower earnings capacity, because native endowments must first be realized through effort in education, which belongs to the private sphere of individual decisions.[9] Hence, in our view, the distinction between the educational and the working stage is key in the Theory and, to the extent that individual responsible decisions in the educational stage matter, the social contract is not to be intended as merely redistributive but, also, the mechanism-design by which incentives to effort in education are determined.

According to Rawlsian *background procedural justice*, if the two Principles above were rigorously implemented in a society, then existing disparities in the ex-post income distribution would be (politically) just. This is because observed income disparities in the labor market would originate exclusively (i) from different effort levels exerted in the labor market, and/or (ii) from wage rate disparities mitigated by

[8] "[T]*he progressive principle of taxation might not be applied to wealth and income for the purposes of raising funds (releasing resources to government), but solely to prevent accumulations of wealth that are judged to be inimical to background justice, for example, to the fair value of the political liberties and to fair equality of opportunity. It is possible that there need be no progressive income taxation at all*" (Rawls 2001, p.161).

[9] "[E]*ven supposing that the least-advantaged.. include many individuals born into the least-favored social class of origin, and many of the least (naturally) endowed and many who experience more bad luck and misfortune, nevertheless those attributes do not define the least advantaged. Rather, it happens that there may be a tendency for such features to characterize many who belong to that group*" (Rawls 2001, p.59).

the scheme of wages (defined behind the veil of ignorance). However, as far as (i) inequalities originating from different effort exerted in the labor market are regarded as fair, and (ii) inequalities obtained through the application of the scheme of wages are designed in such a way as to satisfy the Difference Principle, then the sole just (or '*not unjust*') inequalities would survive. Most importantly, within this theoretical framework, incentives to effort in education and in the labor market would be both preserved even if the sole inequalities that are beneficial to the least-advantaged are permitted.

As far as incentives and efficiency issues are accounted for, it is evident itself that Rawls' contribution goes well beyond opportunity egalitarianism, at least as intended in main approaches to distributive justice. Even though Rawls and Sen are often indicated as the two scholars bringing equality of opportunity back to the attention of the economic literature Roemer and Trannoy (2016), as observed by Sugden (1993), their contributions strongly differ from each other. Sen proposes a theory of 'social good' where capabilities—intended as opportunities to achieve functionings (i.e. states of being)—are identified as a better starting-gate for the measurement of well-being[10]; accordingly, distributive justice is defined in terms of equality of opportunities (or chances) to attain states of being and it is captured by equality of individual capability sets. Opportunities (and liberties) are key in Rawls' theory as well, but his proposal consists of a Theory of Justice, not a theory of social goodness; here, equality of basic liberties (Liberty Principle) and equality of chances to attain offices and positions (Fair Equality of Opportunity) are priorities required "*to regulate social and economic inequalities*" (Rawls 2001, p.42) according to the Difference Principle, which is aimed at granting the stability of political institutions, independently from the maximization of the social good.

3 The model

In this Section, we propose a simple analytical formalization of the Difference Principle by which Rawls' ideal is shown to be time consistent when information is assumed to be progressively revealed in three stages, namely (i) the original position, (ii) the educational stage, and (iii) the working stage. More precisely, the Difference Principle is modeled once Fair Equality of Opportunity is attained, meaning that, educational opportunities are assumed to have already been equalized for all individuals, so that the impact of unequal social class of origin can be totally disregarded for our purposes.

Let $(\theta_i, \theta_j) \in \Re^+$ be the native talent of the *i*th and the *j*th (group of) individuals respectively. According to Rawls theory, native endowments can be inferred ex-post only, in that they are merely potential and cannot come to fruition apart from social conditions and institutions revealing ex-post only. This means that, in a population of two individuals, the native talent—respectively θ_H and θ_L with $\theta_H > \theta_L$—can be

[10] According to both Rawls and Sen, the 'revealed preference welfarism' is a non-starting for a theory of well-being because choices are not necessarily motivated by the search for goodness and, also, provide poor information on it.

only inferred from the earnings capacity observed in their respective job activities (working stage).

Given native talent θ, let Θ be the earnings capacity where, for the sake of simplicity, we assume

$$\begin{aligned} \Theta_i &= e_i\theta_i \\ \Theta_j &= e_j\theta_j \end{aligned} \tag{1}$$

with $e \in \mathfrak{R}^+$ indicating effort in education.[11] The earnings capacity, Θ, is assumed to indicate the money-value of the realized talent obtained at the working age, which might be thought as individual productivity determined by the complementarity between effort in education, e, and native talent, θ. As far as the contribution of θ is inferred ex-post from (1), native talent is defined as the money-value of one unit of effort in education (contingent on social conditions and institutions revealing in the working stage).

Notice that two different states of the world may occur; either $\theta_i = \theta_H$ and $\theta_j = \theta_L$, or $\theta_i = \theta_L$ and $\theta_j = \theta_H$. In addition, as we observed in the previous Section, behind the veil of ignorance the probability is intended in classical terms, so that the two states above are considered as equally probable.[12]

Let w be the individual wage income defined as the amount of primary goods an individual may potentially have access to. In Rawls' Theory, wage income, w, is intended as a potential (not effective) value which is defined up to the entire (or fixed) time endowment, equal for all individuals by definition; formally,

$$\begin{aligned} w_i &= \Theta_i T \\ w_j &= \Theta_j T \end{aligned} \tag{2}$$

with T indicating time endowment. Notably, 'potential wage income' is not to be confused with 'effective wage income', whose size will be determined by the effort exerted in the labor market and which may clearly differ across individuals. For simplicity, we normalize the time endowment, $T = 1$, so that w_i measures both 'potential wage income' and 'wage rate' of the ith individual.

In the absence of any redistributive scheme, the wage rate in the labor market, w, corresponds to the earnings capacity, Θ, so that $w_i = \Theta_i$ and $w_j = \Theta_j$. Instead, if the scheme of wages agreed behind the veil of ignorance applies according to Rawlsian fair system of cooperation, then the wage rate of each individual is not anchored any longer to the corresponding earnings capacity. Specifically, we restrict our focus to the case of linear redistributive systems, so that

[11] The assumption of an equal technology of production for both individuals (1) is not so demanding in the Ralwsian framework, since equal access to education is granted by the implementation of the Fair Equality of Opportunity principle, which is also expected to mitigate the impact of (non-)financial investments in education.

[12] In our formalization of the Difference Principle, the equal probability assumption is imposed to preserve consistency with Rawlsian spirit, however all of the results discussed in this paper do not rely on this assumption. Also, it is worth observing that this assumption is not required for impartiality, whereas the latter is crucial in Harsanyi's (1953, 1955) impartial observer theorem (e.g. Moehler (2015)).

$$\begin{aligned} w_i &= \alpha + (1-\beta)\,\Theta_i \\ w_j &= \alpha + (1-\beta)\,\Theta_j \end{aligned} \tag{3}$$

where $\alpha > 0$ and $\beta \in (0,1)$ identifies the redistributive parameter of the scheme of wages. Remarkably, to the extent that the budget constraint is required to hold, i.e. $(w_i + w_j)T = (\Theta_i + \Theta_j)T$, it must be the case that $\alpha = (\beta/2)(\Theta_i + \Theta_j)$, so that (3) can be equivalently rewritten as

$$\begin{aligned} w_i &= \frac{\beta}{2}\Theta_j + \left(1 - \frac{\beta}{2}\right)\Theta_i \\ w_j &= \frac{\beta}{2}\Theta_i + \left(1 - \frac{\beta}{2}\right)\Theta_j \end{aligned} \tag{4}$$

where the higher is β, the greater is the contribution to the ith wage rate of the jth earnings capacity, and vice versa. Evidently, if $\beta = 0$ then the wage rate corresponds to the earnings capacity of each individual, whereas if $\beta = 1$ then the wage rate of each individual is equally distributed and equally determined by the earnings capacity of all members of the society.

Also, it is worth observing that the redistribution originating from the scheme of wages is order-preserving by construction; indeed, provided that $(w_i - w_j) = (1-\beta)(\Theta_i - \Theta_j)$ with $\beta \in (0,1)$, the following equivalence holds true $w_i \gtreqless w_j \Leftrightarrow \Theta_i \gtreqless \Theta_j$. Hence, the redistributive scheme can only render the least-advantaged individual ‘less’ least-advantaged but can never switch the positions of the least- and the most-disadvantaged individuals.

However, it may happen that the least-advantaged individual in terms of earnings capacity, who is also the least-advantaged in terms of wage rate (and potential wage income), does not coincide with the poorest individual in terms of effective wage income. This is going to happen whether the gap in terms of earnings capacity—which is mitigated by the application of the wage scheme—is more than compensated by a sensibly greater effort exerted in the labor market by the least-advantaged individual. In this special case, the scheme of wages would be redistributing in favor of the richer individual in terms of effective wage income, since this individual is the least-advantaged in terms of earnings capacity.

Given the key definitions of wage rate in (4) and earnings capacity in (1), the timing of the game is crucial for our formalization of the Difference Principle. At time 0 (original position), the two (groups of) individuals define the scheme of wages, which redistributes the overall earnings capacity according to (4) when no information is available on native talents (native traits) and on preference types (i.e. disutility of effort). At time 1 (educational stage), the preference type is associated to each individual and publicly revealed; at this stage, each individual is supposed to choose effort in education in such a way as to maximize his own *expected* utility, given the scheme of wages signed at the previous stage. Once again, the earnings capacity at this stage is contingent on native talents which, according to Rawls, can be inferred at the working age only. At time 2 (working stage), the earnings capacity (productivity) of each individual is publicly observable, so that native talents can be automatically inferred; it is only at the working stage that individuals can really apprehend the productivity of the native talent they have been endowed with (either θ_H or θ_L). Remarkably, this implies that the identity of the least-advantaged individual becomes observable when effort in education has been exerted already.

Hence, the optimal scheme of wages can be defined by backward induction, in that the optimal social contract agreed at time 0 is expected to account for individual decisions on effort in education at stage 1 which, in turn, account for expectations on native talent revealing at time 2. Notably, as far as preference types and native talents become publicly observable at different stages, it must be the case that individual responsibility for effort in education is naturally embedded in the definition of the optimal amount of redistribution behind the veil of ignorance; in addition, it is worth observing that, in line with Rawls' idea, preference types and native talents are not *playing at the same level* when determining the optimal social contract due to progressive revelation of information.

3.1 Educational stage

In this Section, we assume that individuals act rationally by choosing effort in education in such a way as to maximize their objective function, as defined in terms of expected utility. In contrast with the tradition of welfare-consequentialism, the notion of utility is merely indicative in this framework, as it is intended to measure command over resources, that is, the instrumental value of primary goods "*that are generally necessary to enable citizens adequately to develop and fully exercise... their determinate conceptions of the good*" (Rawls 2001, p.57). Evidently, this is not to be confused with the intrinsic value of goods (e.g., happiness, betterness) that characterizes the utilitarian tradition.

We consider a quasi-linear Bernoulli utility function, $U(w, e) = aw + (1 - a)(1 - e^2)$, which depends on (i) contingent wage income, w, and (ii) the dis-utility[13] of effort in education, e. Notably, quasi-linearity implies risk-neutrality of the two individuals which is not jeopardizing Rawlsian spirit of the model.[14]

Let $(a_i, a_j) \in (0, 1)$ indicate the *preference type* of the two individuals which depends on the relative contribution of wage income, w, to overall utility, or, alternatively, on "propensity to effort in education". We assume that the two individuals differ with respect to their preference type with $a_H > a_L$. To simplify the formalization, we hypothesize $a_i = a_H$ (so, $a_j = a_L$), since the opposite case implies perfectly symmetric solutions.

From (1) and (4), we indicate by w_{iH} and w_{iL} the two state-contingent wage incomes of the ith individual for, respectively, $\theta_i = \theta_H$ and $\theta_i = \theta_L$; e.g., w_{iH} is the wage income of the ith individual endowed with preference type a_H, would

[13] Alternatively, the dis-utility may be formalized as a resource cost for education included in the budget constraint of the utility maximization (Phelps 1973).

[14] "*The widespread idea that the argument for the difference principle depends on extreme aversion to uncertainty is a mistake..*" (Rawls 2001, p.43).

native talent θ_H emerge at time 2. Hence, under the quasi-linearity assumption, the expected utilities of the two individuals can be defined as follows

$$\begin{aligned} E[U_i(\cdot)] &= \frac{1}{2}U_i(w_{iL}, e_i) + \frac{1}{2}U_i(w_{iH}, e_i) = a_i\left[E(w_i)\right] + (1-a_i)\left(1-e_i^2\right) \\ E[U_j(\cdot)] &= \frac{1}{2}U_j(w_{jL}, e_j) + \frac{1}{2}U_j(w_{jH}, e_j) = a_j\left[E(w_j)\right] + (1-a_j)\left(1-e_j^2\right) \end{aligned} \tag{5}$$

where e^2 is the disutility from effort, which is assumed to be increasing and convex, whereas $E(w)$ is the expected wage income defined as

$$\begin{aligned} E(w_i) &= \frac{1}{2}w_{iH}(e_i, e_j, \theta_H, \theta_L, \beta) + \frac{1}{2}w_{iL}(e_i, e_j, \theta_H, \theta_L, \beta) \\ E(w_j) &= \frac{1}{2}w_{jL}(e_i, e_j, \theta_H, \theta_L, \beta) + \frac{1}{2}w_{jH}(e_i, e_j, \theta_H, \theta_L, \beta) \end{aligned} \tag{6}$$

where $w_{ik}(\cdot), w_{jk}(\cdot)$, $k = H, L$, are, respectively, the *i*th and the *j*th state-contingent wage incomes.

It is worth observing that quasi-linearity of preferences, together with linearity of contracts (β), offsets strategic interactions due to the interdependence among individual utilities. This allows to keep the model as simple as possible, while preserving the focus on the impact of redistribution on optimal effort decisions.[15]

The optimal decisions of effort in education associated to each propensity to effort, respectively to the *i*th and to the *j*th individual, are

$$\begin{aligned} e_H^* &= \frac{a_H(2-\beta)(\theta_H + \theta_L)}{8(1-a_H)} \\ e_L^* &= \frac{a_L(2-\beta)(\theta_H + \theta_L)}{8(1-a_L)} \end{aligned} \tag{7}$$

Hence, the greater is β, the lower is the incentive of each individual to exert effort to acquire earnings capacity. Formally, any increase in β reduces the impact of the earnings capacity acquired by an individual on his own wage rate, while it increases the influence of the earnings capacity of the rest of the population. As such, a standard free-rider effect arises (Holmstrom 1982) because, as far as the social contract is agreed, effort in education of each individual is intended to contribute to the production of a *social output*—the overall amount of earnings capacity produced in a society—which is to be shared among the population according to the scheme of wages agreed at the previous stage, i.e. behind the veil of ignorance.

It is also worth observing that the individual with a higher propensity to effort will always opt for a greater effort at time 1. In particular, as far as the individual with the higher propensity to effort at time 1 does not necessarily correspond to the individual with the greater wage income at time 2, there is no reason for the better

[15] Alternatively, strategic interactions might be allowed with quasi-linear preferences by considering the share, instead of the level, of earnings capacity of each individual. However, while enriching the basic framework, this alternative would not alter any of the main results in Sect. 4.

preference type at time 1 to conceal (or, unreveal) its propensity to effort, i.e. to miss the opportunity to realize its own native talent at time 1 (Rawls 1974).

4 Original position

Given the optimal effort each preference type is willing to exert, the optimal scheme of wages, β, can be defined by solving backward, i.e. behind the veil of ignorance. In what follows, we will refer to β^* as the optimal social contract even if, as we said above, the redistributive parameter is, more generally, the output of the two Principles (Liberty and Equality) agreed in the social contract.

Let the *i*th individual be the one endowed with the higher propensity to effort, so that $a_i = a_H$. Since the two individuals differ from each other in terms of native talents, i.e. θ_H and θ_L with $\theta_H > \theta_L$, two different states of the world are to be considered: either (i) the native talent of the *i*th individual (with higher propensity to effort) reveals of type θ_H (implying *j*'s θ_L-type), which we refer to as 'concordant-state', or (ii) the native talent of the *i*th individual reveals of type θ_L (implying *j*'s θ_H -type), which we refer to as 'discordant-state'.

Remarkably, in the concordant-state the *i*th individual corresponds to the "most-advantaged", whereas *j* is the "least-advantaged". Differently, in the discordant-state, the least-advantaged cannot be identified *a priori* since the individual with a better propensity to effort is the penalized one in terms of native talent, and vice versa.

In the concordant-state, let Θ_{HH} (resp. Θ_{LL}) be the earnings capacity of the *i*th (resp. *j*th) individual with higher (resp. lower) propensity to effort and better (resp. worse) native talent, as obtained by replacing in (1) the optimal effort decisions from (7) with $e_i^* = e_H^*$, $e_j^* = e_L^*$, $\theta_i = \theta_H$, $\theta_j = \theta_L$. Clearly, $\Theta_{HH} > \Theta_{LL}$. As such, let w_{HH} and w_{LL} be the *state-contingent* (potential) wage incomes (or, wage rates) obtained from Θ_{HH} and Θ_{LL} by implementing the scheme of wages in (4) where, as observed in the previous Section, $\Theta_{HH} > \Theta_{LL}$ implies $w_{HH} > w_{LL}$ (and vice versa). According to Rawls, if the concordant-state occurs, then the least-advantaged individual is the LL-type, i.e. the individual with the worst endowment in terms of both talent and propensity to effort.

In the discordant-state, let Θ_{HL} (resp. Θ_{LH}) be the earnings capacity of the *i*th (resp. *j*th) individual with higher (resp. lower) propensity to effort and worse (resp. better) native talent, as obtained by replacing in (1) the optimal effort from (7) with $e_i^* = e_H^*$, $e_j^* = e_L^*$, $\theta_i^* = \theta_L^*$, $\theta_j = \theta_H$. Also, let w_{HL} and w_{LH} be the *state-contingent* (potential) wage incomes obtained from Θ_{HL} and Θ_{LH} as before. Here, the least-advantaged individual may be either the one endowed with the lower propensity to effort but better native talent (i.e. $\Theta_{HL} > \Theta_{LH}$), or the one endowed with higher propensity to effort but worse native talent (i.e. $\Theta_{HL} < \Theta_{LH}$). Formally, in the discordant-state, the least-advantaged individual is identified by the following (equivalence) condition.

$$\Theta_{HL} \gtreqless \Theta_{LH} \iff a_H(1 - a_L)\theta_L \gtreqless a_L(1 - a_H)\theta_H \tag{8}$$

The equivalence condition in (8) 'defines' the least-advantaged position in the discordant-state but, most importantly, it does not allow to 'identify' the least-advantaged individual behind the veil of ignorance (when preference types and native abilities are not observable yet), nor does it allow such identification at the education stage (when native abilities have not revealed yet). Hence, since the identity of the least-advantaged may be realized at the working stage only, it must be the case that, at the educational stage, none of the two individuals is willing to replace income (consumption) with leisure to avoid the burden of redistribution; there is no incentive to false revelation of the preference type at the educational stage.[16]

Since two different and equally probable states—concordant and discordant—must be accounted for, the β^* that maximizes the wage rate, w, of the least-advantaged is inevitably state-contingent.

In Section 4.1, the two *state-contingent* optimal social contracts, i.e. for the concordant and discordant-state, are determined separately; each of them implies a potential wage income distribution at time 2. In Section 4.2, given the two state-contingent distributions of potential wage incomes (hereafter, wage incomes), the optimal social contract, β^*, is determined under uncertainty conditions, which is done by evoking the notion of universally ex-post efficiency (Starr 1973; Harris 1978; Hammond 1981).

4.1 State-contingent optimal contracts

According to the definition of 'ex-post k-efficiency' (Harris 1978), an allocation is said to be efficient in state k if there is no feasible allocation such that, *in state* k, the utility of an individual is increased without worsening the utility of another individual. This notion of 'ex-post k-efficiency' can be re-adapted within Rawlsian framework by applying the same rule to the sole wage income, w, of the least-advantaged. Specifically, according to the Difference Principle, *for each state*, the two optimal (state-contingent) contracts, β_1^* and β_2^*, are determined by maximizing, respectively, the wage income, w, of the least-advantaged individual in the concordant-state, i.e. w_{LL}, and in the discordant-state, i.e., either w_{LH} or w_{HL} depending on condition (8).

It is worth observing that any variation of the scheme of wages, β, generates two different effects on wage incomes. On the one hand, according to (4), any increase of β implies a redistribution in terms of earnings capacity from the most to the least-advantaged type, meaning that β is a redistributive parameter (direct effect). On the other hand, β acts as a sort of wage-premium determining the dis-incentive to effort; specifically, from (7), if β increases then the relative contribution of the ith (jth) earnings capacity to its own earnings capacity decreases, so that any individual is less willing to make high effort in education (indirect effect). In this sense, a dis-incentive effect is to be considered too.

Evidently, the redistributive and the dis-incentive effect are both reducing the wage income of the most-advantaged, whereas a trade-off occurs in the case of the least-advantaged individual. For the latter, if the dis-incentive effect is dominating

[16] The possibility of a leisure trade-off is the core of Musgrave's critique to the Theory (Musgrave 1974). Our model seems to provide a formal reply to Musgrave's critique as well.

for all β's, then the wage income, w, is strictly decreasing in β, so that redistribution is never desirable ($\beta^* = 0$). Differently, if the redistributive effect dominates, for some β at least, the dis-incentive effect of the least-advantaged, then the wage income, w, is increasing in β over this range, so that redistribution is desirable ($\beta^* > 0$). In addition, the redistributive effect becomes more and more important when the gap between individual earnings capacities increases. These aspects are formalized for each state (i.e., concordant and discordant) in the two following Propositions.

Proposition 1 ***(State-contingent optimality in concordant-state)*** *Let* $\Delta\Theta_1 = (\Theta_{HH} - \Theta_{LL}) > 0$ *be the earnings capacity gap in the concordant-state, there exists* $k_1 > 0$ *such that*

- $\forall\ a_L, a_H, \theta_L, \theta_H : \Delta\Theta_1 \leq k_1$, wage income of the least-advantaged is strictly decreasing with respect to $\beta \in (0, 1)$, therefore $\beta_1^* = 0$;
- $\forall\ a_L, a_H, \theta_L, \theta_H : \Delta\Theta_1 > k_1$, wage income of the least-advantaged is inverse U-shaped with respect to $\beta \in (0, 1)$, and

$$\beta_1^* = \frac{a_H(a_L - 1)\theta_H - 2a_H a_L \theta_L + 2a_L\theta_L}{a_H(a_L - 1)\theta_H - a_H a_L \theta_L + a_L\theta_L}$$

Proof See Appendix A.1. □

Differently, in the discordant-state, two different possibilities must be considered because the individual with the higher propensity to effort and worse native talent might be either the most- or the least-advantaged depending on condition (8). Specifically, we denote by β_{21}^* the optimal (state-contingent) contract whether the least-advantaged individual corresponds to the one endowed with better native talent, and by β_{22}^* the optimal (state-contingent) contract whether the least-advantaged is the individual with worse native talent.

Proposition 2 (***State-contingent optimality in discordant-state***) *Let* $\Delta\Theta_2 = (\Theta_{HL} - \Theta_{LH})$ *be the earnings capacity gap in the discordant-state. If* $\Delta\Theta_2 > 0$, *then* $\exists k_{21} > 0$:

- $\forall\ a_L, a_H, \theta_L, \theta_H : \Delta\Theta_2 \leq k_{21}$, wage income of the least-advantaged is strictly decreasing with respect to $\beta \in (0, 1)$, therefore $\beta_{21}^* = 0$;
- $\forall\ a_L, a_H, \theta_L, \theta_H : \Delta\Theta_2 > k_{21}$, wage income of the least-advantaged is inverse U-shaped in $\beta \in (0, 1)$, and

$$\beta_{21}^* = \frac{2(a_H - 1)a_L\theta_H - a_H a_L \theta_L + a_H\theta_L}{(a_H - 1)a_L\theta_H - a_H a_L \theta_L + a_H\theta_L}$$

If $\Delta\Theta_2 < 0$,[17] *then* $\exists k_{21} > 0$:

- $\forall\ a_L, a_H, \theta_L, \theta_H : (-\Delta\Theta_2) \leq k_{22}$, wage income of the least-advantaged is strictly decreasing in $\beta \in (0,1)$, therefore $\beta^*_{22} = 0$;
- $\forall\ a_L, a_H, \theta_L, \theta_H : (-\Delta\Theta_2) > k_{22}$, wage income of the least-advantaged is inverse U-shaped in $\beta \in (0,1)$, and

$$\beta^*_{22} = \frac{(a_H - 1)a_L\theta_H - 2a_H a_L\theta_L + 2a_H\theta_L}{(a_H - 1)a_L\theta_H - a_H a_L\theta_L + a_H\theta_L}$$

Proof See Appendix A.2. □

Propositions 1 and 2 show that, for redistribution to be desirable behind the veil of ignorance (i.e., $\beta^* > 0$), wage income of the least-advantaged must be inverse U-shaped which is going to be the case when the earnings capacity gap is not small enough.

Basically, when wage income of the least-advantaged is inverse U-shaped, redistribution is desirable at $\beta = 0$, so that β is increased. However, when β increases, the dis-incentive effect becomes stronger for both the least- and the most-advantaged individuals, so that the cake to be redistributed is reduced, and the redistributive effect jeopardized; evidently, the optimal state-contingent contract is obtained when the dis-incentive and the redistributive effects perfectly compensate to each other at the margin for the least-advantaged.

Most importantly, it is worth observing that, once the optimal (state-contingent) social contract has been achieved, any additional increase in redistribution would not ameliorate the distribution of wage incomes, proving that legitimate inequalities are clearly permitted in the Theory of Justice as Fairness.

On the other way around, this explains why, starting from a perfectly egalitarian social contract ($\beta = 1$), any marginal increase of inequality (i.e., diminishing β) induces higher effort of both individuals in such a way as to enhance their wage incomes; for the least-advantaged, the effect of the smaller redistribution is initially more than compensated by the increasing cake due to greater incentive to effort for both, the most and the least-advantaged individual. As such, a marginal decrease of β starting from $\beta = 1$ generates Pareto improvements (and so, economic growth). Subsequently, once the break-even point is achieved, for any additional increase of inequality, the redistributive effect becomes dominating for the least-advantaged individual, so that its wage income decreases. From now on, any additional increase of inequality—even if growth enhancing — is not bought by the least-advantaged individual in that, growth is not of the pro-poor kind.

[17] Notice that, if $a_H(1 - a_L)\theta_L = a_L(1 - a_H)\theta_H$, then the two individuals are equally endowed in terms of earnings capacity (8), so that there is no inequality in terms of wage income; as far as the sole inequalities in terms of the (endogenous) realized, not native, talent matter, this case is irrelevant.

Finally, from the comparison between optimality conditions in Propositions 1 and 2, it turns out that redistribution is greater in the concordant-state as compared to the discordant case. This is formalized in the following Corollary.

Corollary 1 $\beta_1^* > \beta_{21}^*$ *and* $\beta_1^* > \beta_{22}^* \; \forall \; a_L < a_H$ *and* $\theta_L < \theta_H$.

Proof See Appendix A.3. □

Basically, the size of the earnings capacity gap, $\Delta\Theta$, is shown to matter both, within the same state (Propositions 1 and 2), as well as across different states (Corollary 1); specifically, the greater is the earnings capacity gap originating from endowments (preferences and native abilities), the more redistribution is expected to characterize the social contract. In addition, from Corollary 1, it must be the case that the size of the redistribution operated by the optimal contract (β) is increasing in the native talent gap, provided that the least-advantaged is endowed with the lower native talent (and vice versa).

4.2 Optimal contract under uncertainty conditions

In the previous Section, two state-contingent optimal contracts, for the concordant and the discordant-state respectively, have been identified. Specifically, it can be shown that, for any $a_H, a_L, \theta_H, \theta_L \in (0, 1)$, (i) $\beta_1^* = \beta_{21}^*$ if and only if $\theta_H = \theta_L$, whereas (ii) $\beta_1^* = \beta_{22}^*$ if and only if $a_H = a_L$. Intuitively, to the extent that β_1^* and β_{21}^* are both obtained when the least-advantaged corresponds to the individual with the lowest propensity to effort, it must be the case that the difference can be originating from the native talent gap only. Similarly, when considering β_1^* and β_{22}^*, the least-advantaged is characterized by the worse native talent, but different preferences.

As such, unless valid motivations are adduced by which one or the other state is neglected on *a priori* grounds, the optimal contract(s), which we denote by β^*, is to be defined under uncertainty conditions.

According to the existing literature (Starr 1973; Harris 1978; Hammond 1981), different approaches can be used to define efficiency under uncertainty conditions. Even if the debate between different optimality conditions in the presence of uncertainty conditions is not the object of our analysis, let's recall the distinction made between 'ex-ante efficiency' and 'universal ex-post efficiency'.

By the former, an allocation is said to be ex-ante efficient if there is no feasible allocation so that the expected utility (e.g., von Neumann-Morgenstern) of an individual can be enhanced without worsening the expected utility of another individual. Differently, by the latter, an allocation is said to be universally ex-post efficient if there is no feasible allocation such that, *for each possible state*, the utility of an individual is increased without worsening the utility of another individual.

Consequently, by virtue of ex-ante efficiency, an 'ex-ante Pareto improvement' occurs if all individuals are indifferent, and at least one individual strictly prefers allocation x as compared to y in terms of expected utility. Instead, an 'universal ex-post Pareto improvement' is obtained when all individuals are indifferent in each state, and at least one individual in one state is better-off in x as compared to y. Evidently, the universal ex-post approach is much more demanding than the ex-ante approach; however, the universal ex-post approach is the only one ensuring ex-post consistency of efficiency orderings, meaning that, if an allocation is strictly preferred under uncertainty conditions, then the same allocation is still preferred once the information has revealed.

Coming back to our model, to the extent that both individuals have access to the same (empty) information set at time 0 (i.e., behind the veil of ignorance), the 'ex-ante efficiency' approach would be a non-starting for egalitarianism, as both individuals would be clearly associated to the same expected wage income, w, as defined with respect to the four equally-probable and mutually-exclusive possible states (i.e., w_{HH}, w_{HL},w_{LH}, w_{LL}).

The universal ex-post approach is definitely to be preferred for our purposes. By the latter, (state-contingent) wage incomes are not aggregated across different states at the individual level. Instead, an ordering among different schemes is defined by comparing state-contingent distributions of wage incomes with different degrees of inequality, which is the very scope of the Rawlsian Difference Principle.[18]

In what follows, universal ex-post efficiency is implemented to characterize the optimal contract(s) under uncertainty conditions. Consistently with the Rawlsian framework, dominance conditions are applied to the distribution of wage incomes, not utilities.[19] Two different formalization of universal ex-post efficiency are considered. First, we implement the standard idea of 'universal ex-post Pareto-efficiency', by which optimality is defined by accounting for the wage income of both, the most- and the least-advantaged individual in the concordant and discordant-state. Next, since the bulk of the Theory of Justice as Fairness is aimed at legitimating the sole inequalities which are improving the condition of the least-advantaged individual (maximin), universal ex-post Rawls-efficiency' is defined by focusing exclusively on the least-advantaged individual in the two states. Basically, Pareto orderings rely on unanimity judgments on ex-post conditions (independently from distributive judgments), whereas Rawlsian orderings require an unanimous agreement in the original position only (while implementing a distributive judgment).

As it will be clearer in what follows, when moving from certainty to uncertainty conditions, universal ex-post Pareto-efficiency does not alter the nature of the Pareto dominance criterion, which is a *partial ordering* independently from uncertainty. On the contrary, the introduction of uncertainty sensibly modifies the Rawls criterion, which is a *complete ordering* under certainty conditions (in that β is uniquely defined in each state), but a *partial ordering* when uncertainty is accounted for. Most importantly, the set of optimal contracts, as obtained in terms of universal ex-post Rawls-efficiency, is shown to be a subset

[18] "*Consider a situation in which an impending climate change will alter the distribution of well-being on Earth. Suppose that only two scenarios are considered possible. In one scenario, the extreme latitudes gain and the low latitudes suffer, whereas the reverse occurs in the other scenario*" (Fleurbaey 2010). Even if ex-post egalitarianism is inevitably jeopardized, the same climate change would be harmless in terms of expected utilities.

[19] Under standard symmetry assumptions, if the utility (increasing) of each individual depends on its wage income only, then Pareto efficiency is equivalently defined with respect to the distributions of incomes and utilities (Amiel and Cowell 1994).

of the universally ex-post Pareto-optimal contracts, which makes the notion of Rawls-efficiency more ‘discriminating’ for policy purposes.

4.2.1 Universal ex-post Pareto-efficiency

From the previous Section, let’s consider the relationship between the wage incomes of the two individuals, w_i and w_j, for all possible values of $\beta \in (0,1)$ in each of the two states, which we will refer to as ‘state-contingent Rawls-efficiency frontiers’. Specifically, as concerns the discordant-state, we only consider the case in which the jth individual is the least-advantaged, whose corresponding optimal contract is β^*_{21}. This allows for a better and more immediate understanding of optimality conditions under uncertainty conditions.

First, it is worth observing that, since the wage income of the most-advantaged is strictly decreasing with respect to β (redistributive and dis-incentive effects move in the same direction), the wage income of the least-advantaged can be simultaneously plotted with respect to β and the wage income of the most-advantaged individual (see Appendix A.4). This is done in Fig. 1, where β is *decreasing* along the x-axis by construction. Clearly, if the wage income of the least-advantaged individual is inverse U-shaped with respect to β, then the (state-contingent) Rawls-efficiency frontier must be inverse U-shaped as well.[20] Differently, if the wage income of the least-advantaged is strictly decreasing with β (i.e., in the case of a sufficiently small earnings capacity gap, $\Delta\Theta$), then the corresponding frontier must be positively sloped (Fig. 1). More precisely, provided that $\beta^*_1 \geq \beta^*_{21} \geq 0$, if $\beta^*_1 = 0$ then $\beta^*_{21} = 0$, not vice versa; equivalently, if the wage income of the least-advantaged in the concordant-state is strictly decreasing with respect to β, then it must be strictly decreasing in the discordant-state as well.

Since the sole index of primary goods—(potential) wage income — matters in the Rawlsian framework, rank-dominance criteria apply (Saposnik 1981; Amiel and Cowell 1994), so that universal ex-post Pareto-efficiency is defined as follows.

Definition 1 ***(Universal ex-post Pareto-efficiency)*** An allocation is said to be universally ex-post Pareto-optimal if there is no other feasible allocation by which the wage income of one individual cannot be increased without worsening the wage income of the other individual both in the concordant and in the discordant-state.

Formally, let $\beta^A \in (0,1)$ be the contract whose corresponding state-contingent distributions of wage incomes are $\bar{w}^A_1 = \{w^A_{HH}, w^A_{LL}\} = \{w^A_{LL}, w^A_{HH}\}$ (concordant-states), and $\bar{w}^A_2 = \{w^A_{HL}, w^A_{LH}\} = \{w^A_{LH}, w^A_{HL}\}$ (discordant-states), with the equivalence conditions holding by symmetry. As such, we say that β^A is universally ex-post optimal if there is no $\beta^B \neq \beta^A$ such that, together, (i) $\bar{w}^B_1$ is a Pareto improvement of $\bar{w}^A_1$, and (ii) $\bar{w}^B_2$ is a Pareto improvement of $\bar{w}^A_2$.

[20] In Fig. 1, the maximum wage income of the least-advantaged is greater in the discordant case as compared to the concordant one; however this is not necessarily the case as the opposite result may occur as well.

Proposition 3 ***(Universally ex-post Pareto-optimality)*** *The set of universally ex-post Pareto-optimal social contract is*:

- $\beta^* = 0$, if the wage income of the least-advantaged individual is strictly decreasing with respect to β in the concordant-state;
- $0 \leq \beta^* \leq \beta_1^*$, if the wage income of the least-advantaged individual is inverse U-shaped in the concordant-state, whatever the discordant-state.

Proof Straightforward from Appendix A.1, A.2, and A.4. □

Proposition 3 highlights that, if the Rawls-efficiency frontier is strictly increasing in the concordant-state (which implies a strictly increasing frontier in the discordant-state as well), then $\beta^* = 0$; that is, by reducing β (i.e., by moving to the right on the x-axis in Fig. 1), it must be the case that both individuals are made better off, whatever the state, until $\beta^* = 0$ is achieved.

Instead, if the frontier is strictly increasing in the discordant-state only (so, inverse U-shaped in the concordant-state), then, by reducing β, individuals are made better off in both states until β_1^* is achieved; this is sufficient to exclude optimality of the β's in the interval $(\beta_1^*, 1)$. On the contrary, once β_1^* is achieved, by moving further to the right on the x-axis, i.e. increasing the wage income of the most-advantaged, it must be the case that there exists at least one state, that is the concordant-state, by which the least-advantaged individual is made worse off. To the extent that universal ex-post Pareto improvements are not attainable any longer, all β's in $(0, \beta_1^*)$ are universally ex-post optimal.

Finally, if the two Rawls-efficiency frontiers are both inverse U-shaped like in Fig. 1, all social contracts such that $\beta_1^* < \beta < 1$ (left-side in Fig. 1) cannot be optimal in that, as before, the wage income of both individuals can be increased by switching to β_1^*. Instead, for all β's such that $0 < \beta < \beta_1^*$ (right-side in Fig. 1), optimality holds true because there are no alternative schemes by which an universal ex-post Pareto improvement can be obtained; by reducing β from β_1^*, i.e. increasing the wage income of the most-advantaged, there exists at least one state—that is the concordant-state — by which the least-advantaged individual is made worse off.

4.2.2 Universal ex-post Rawls-efficiency

Universal ex-post Pareto-optimality is supposed to account for the (potential) wage income of both, the most- and the least-advantaged individual, in a way that resembles the idea of Pareto-dominance. However, in line with Propositions 1 and 2, the bulk of the Theory of Justice as Fairness is aimed at improving the sole condition of the least-advantaged individual (maximin). In this sense, universal ex-post Pareto-optimality, as defined in Proposition 3, may be weakened according to the maximin principle by focusing exclusively on the least-advantaged individual as follows.

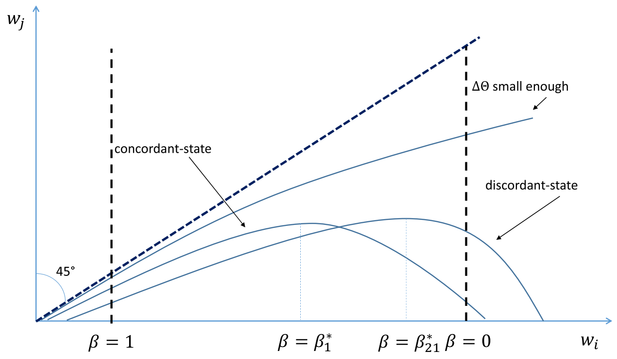

Fig. 1 Rawls-efficiency frontiers

Definition 2 (**Universal ex-post Rawls-efficiency**) An allocation is said to be universally ex-post Rawls-optimal if there is no other feasible allocation by which the wage income of the least-advantaged individual is increased in one state without decreasing in the other state.

In this view, the definition of the optimal social contract becomes less stringent as compared to the standard universal ex-post Pareto-efficiency. The following Proposition identifies, according to Definition 2, the intervals the optimal scheme of wages must belong to, depending on the shape of the (state-contingent) Rawls-efficiency frontier.

Proposition 4 ***(Universally ex-post Rawls-optimality)*** *The set of universally ex-post Rawls-optimal social contract is*:

- $\beta^* = 0$, if the wage income of the least-advantaged individual is strictly decreasing with respect to β in the concordant-state;
- $0 \leq \beta^* \leq \beta_1^*$, if the wage income of the least-advantaged individual is strictly decreasing with respect to β in the discordant-state but inverse U-shaped in the concordant-state;
- $\beta_{21}^* \leq \beta^* \leq \beta_1^*$, if the wage income of the least-advantaged individual is inverse U-shaped in both the concordant and the discordant-state.

Proof Straightforward from Appendix A.1, A.2, and A.4. □

Although the $\beta^* = 0$ solution is the same as in Proposition 3, the second solution (i.e., $\beta^* > 0$) is now more articulated in that, two different scenarios are to be considered. More precisely, if the Rawls-efficiency frontier is inverse U-shaped in the concordant-state, and strictly increasing in the discordant-state, then it must be the case that all contracts such that $\beta \in [\beta_1^*, 1)$ can be ameliorated according to Definition 2 by opting for β_1^*. Moving further to the right from $\beta = \beta_1^*$, to the extent that the frontier is inverse U-shaped in the concordant-state, there is no alternative contract by which the wage income of the least-advantaged is increased independently from the state; this is similar to the result obtained in Proposition 3.

If both frontiers are inverse U-shaped, then contracts in the interval $\beta \in [\beta_1^*, 1)$, as before, cannot be optimal. However, in contrast with Proposition 3, the rest of the contracts are not necessarily optimal any longer because, for all contracts such that $\beta > \beta_{21}^*$, the wage income of the least-advantaged increases independently from the state. Consequentially, the sole contracts such that $\beta \in (\beta_{21}^*, \beta_1^*)$ are universally ex-post Rawls-optimal. Evidently, as compared to Proposition 3, universally ex-post Rawls-optimal social contracts are a subset of the more general universal ex-post case.

From Definition 2, *partial justice orderings*[21] can be derived accordingly. Formally, let $w_{LL}^j(B)$, $w_{LH}^j(B)$, $w_{LL}^j(A)$ and $w_{LH}^j(A)$ be the state-contingent (potential) wage income of the (jth) least-advantaged individual as obtained when the contracts β^A and β^B are considered, with the subscripts LL and LH referring to the concordant and the discordant-state respectively. Also, let $\beta^B \succ \beta^A$ indicate that β^B is strictly preferred to β^A, with $\sim$ indicating the symmetric component of the *justice ordering*, whereas $\beta^B || \beta^A$ signifies that β^B and β^A are non-comparable.

According to Definition 2,

$$w_{LL}^j(B) \gtrless w_{LL}^j(A),\ w_{LH}^j(B) \gtrless w_{LH}^j(A) \Longleftrightarrow \beta^B \gtrless \beta^A;\ \beta^B || \beta^A \ \textit{otherwise}.$$

Basically, for an ‘universal ex-post Rawls improvement’ to occur, a contract must be enhancing the wage income of the least-advantaged in both, the concordant and the discordant-state.

Two observations are required concerning, respectively, the relation between Rawls improvements and Rawls-optimality, as well as the comparison between Rawls and Pareto improvements. First, the optimality of a contract does not imply that this is to be preferred to a non-optimal one; indeed, universal ex-post Rawls-optimality is neither a necessary, nor a sufficient condition for the universal ex-post Rawls improvement to occur.[22]

Second, and most importantly, when considering optimal contracts, universal ex-post Rawls-efficiency is shown to imply universal ex-post Pareto-efficiency, not vice versa. However, as regards universal ex-post Pareto and Rawls improvements, the two criteria are shown to be equivalent if, and only if, the attention is restricted to the sole contracts ensuring economic growth $(\beta_1^*, 1)$; specifically, universal ex-post

[21] Rawls expressly refers to justice orderings, not individual or social welfare ones, where different levels of justice are said to “*represent how claims to goods cooperatively produced are to be shared among those who produced them, and they reflect an idea of reciprocity*” (Rawls 2001, p.62).

[22] Clearly, it is not necessary because $\beta^B \succ \beta^A$ may occur even if $\beta^B, \beta^A \notin (\beta_{21}^*, \beta_1^*)$. In addition, sufficiency does not hold because the optimality of β^B (i.e., $\beta^B \in (\beta_{21}^*, \beta_1^*)$) and the non-optimality of β^A (i.e., $\beta^A \notin (\beta_{21}^*, \beta_1^*)$) do not necessarily imply $\beta^B \succ \beta^A$; e.g., let’s suppose that (i) $\beta^B \in (\beta_{21}^*, \beta_1^*)$ and (ii) $\beta^A \in (0, \beta_{21}^*)$. By (i) and (ii), it must be the case that $w_{LL}^j(B) > w_{LL}^j(A)$, meaning that, in the concordant-state, the potential wage income of the least-advantaged is higher when β^B is implemented. However, if $w_{LH}^j(B) < w_{LH}^j(A)$, then β^A is to be preferred in the discordant-state. To the extent that the two schemes of wages are differently ranked depending from the state, by definition of ‘universal ex-post Rawls improvement’, it must be the case that β^B and β^A are not comparable (i.e., $\beta^B || \beta^A$) in the case above.

Pareto improvements imply economic growth, whereas universal ex-post Rawls improvements might be obtained in the presence of negative growth as well $(0,\beta_2^*)$.[23] This is an immediate consequence of the introduction of a distributive value judgment — the Difference Principle—which is absent in Pareto dominance.

5 Concluding remarks: what's new?

In the existing literature, Rawls' Theory is usually evoked to underpin infinite aversion to inequality in social welfare analysis. Starting from Alexander (1973), the Rawls' *maximin* criterion has been usually represented by Leontief preferences to rank utility distributions originating from a fixed (exogenous) amount of resources (e.g., income).

In this paper, according to Rawls' Theory, the sole inequalities of primary goods, not utility, are considered. Specifically, we refer to the (state-contingent) distributions of potential wage income, with the latter indicating the index of primary goods associated to each individual at the working age. In addition, the potential wage income is assumed to be co-determined by both native talent and effort in education, where the latter is endogenously determined by preference types (or, ambitions) characterizing the propensity to (or, the cost of) effort in education. As such, in our model, preferences capture the instrumental value of potential wage income (i.e. command over resources), and not its intrinsic value in terms of some notion of *betterness* (or *goodness*).

To the extent that the overall time endowment—equal for all by definition—is intended as primary good, in our framework, inequalities in terms of potential wage income are independent from the leisure/effort decision in the labor market, meaning that, in contrast with the old tradition of *welfare-consequentialism* and according to *background procedural justice*, real income distributions are irrelevant within the Rawlsian perspective we propose.

Remarkably, according to our economic formalization of the Difference Principle, we assume that information on effort in education and native talent are progressively revealed over time. With this purpose in mind, we consider a three-stages sequential equilibrium consisting of the original position (or, veil of ignorance), the educational stage, and the working stage. Since the preference type is assumed to be revealed at the educational stage, whereas native talent (as influenced by the shape of social institutions) is quantifiable at the working stage only, it must be the case that, by backward induction, responsibility for individual decisions is automatically accounted for when determining the optimal social contract behind the veil of ignorance. In this sense, the implementation of a three-stage sequential equilibrium allows for a time-consistent interpretation of Rawls thought while preserving a

[23] Quoting Rawls (2001, p.63), "*A further feature of the Difference Principle is that it does not require continual economic growth over generations to maximize upward indefinitely the expectations of the least advantaged (assessed in terms of income and wealth). That would not be a reasonable conception of justice*".

role for individual responsibility—often questioned in the existing literature—with respect to both effort in education and effort in the labor market.

In addition, as far as native talent is inferred ex-post, the identification of the least-advantaged is possible at the working stage only. This implies that, in our model, the least-advantaged does not necessarily correspond to the individual with the worse native talent, because the better endowment in terms of native talent might be more than compensated by the worse endowment in terms of propensity to effort (in education). To the extent that the better (worse) endowed in terms of native talent might be either the better, or the worse endowed in terms of propensity to effort, two different states of the world, i.e., the concordant and the discordant-state, must be accounted for. Therefore, the Difference Principle is modeled under uncertainty conditions according to the definition of 'universal ex-post efficiency'.

In social welfare analysis, the presence of uncertainty conditions is known to characterize Harsanyi's (1953, 1955) impartial observer as well. However, we argue that the two frameworks strongly differ from each other with respect to their ultimate end. Rawls' veil of ignorance is aimed at the definition of an (unanimous) agreement (*social contractualism*) between free and equal persons concerning the identification of legitimate inequalities, whereas Harsanyi's ignorance is used to obtain an impartial definition of social welfare in terms of *betterness*.[24] As such, behind the veil of ignorance, Rawls' souls are supposed to assess inequalities by viewing themselves as potential occupants of each position in a distribution, independently from the identity and preferences of each individual (Saposnik 1981), whereas Harsanyi's "*impersonality requires that the observer have an equal chance of being put in the place of any individual member of the society, with regard not only to his objective social (and economic) conditions, but also to his subjective attitudes and tastes*" (Mongin 2001). Last but not least, to the extent that inequality, not social welfare, is indicated as the object of Rawls' Theory, the uncertainty behind Rawlsian veil of ignorance does not concern exclusively the individual position within a distribution, but, mostly, the possibility of alternative distributions (i.e. concordant and discordant-state) with different degrees of inequality. In this scenario, the notion of expected utility (von Neumann-Morgenstern), which is essential in Harsanyi's Theory, is a non-starting for Rawlsian uncertainty, as it would obscure the inequality of state-contingent distributions which, instead, is captured by the notion of universal ex-post efficiency.

Given the uncertainty conditions above, we draw a separating line between state-contingent and overall optimality of the social contract (respectively, Section 4.1 and Section 4.2). Within each state, the state-contingent contract yields two effects, the dis-incentive and the redistributive effect, which are shown to be conflicting to each other for the least-advantaged individual. As such, for each state (concordant or discordant), redistribution is found to be desirable if and only if there exists at least a state-contingent contract such that the redistributive effect over-compensates the dis-incentive effect for the least-advantaged. If this is the case, then redistribution is desirable until the reduction of the cake—which is induced by the dis-incentive

[24] For details on this distinction see Hampton (1980).

effect—is not so strong to jeopardize the redistributive effect. In this sense, we suggest that Rawls' contribution goes well beyond distributive justice in such a way as to strain into the existing literature on the equity-efficiency trade-off, where the size of the (potential) 'cake' is immediately affected by its distribution (Mirrlees 1971; Phelps 1973; Stiglitz 1987); specifically, in line with Phelps (1973), redistribution is permitted until the income/utility of the least-advantaged individual is maximized.

As a major departure from this literature, in our model the efficiency loss is originating from the free-riding due to the Rawlsian ideal of social cooperation; as far as individuals are assumed to exert an effort to produce a *social output*—the overall earnings capacity—to be shared among members of the community, effort is going to be more and more distorted the stronger is the redistribution operated by the scheme of wages.

In addition, we introduce a dynamic setting where the social contract is defined and unanimously agreed behind the veil of ignorance, whereas optimal effort in education is decided at the next stage, when the native talent has not revealed yet. This difference with the common understanding of the equity-efficiency trade-off is substantial. In our model, the identity of the least-advantaged is unknown at the time of the social contract, so that the maximin principle can be unanimously 'agreed' by both, the most- and the least-advantaged individuals behind the veil of ignorance. Differently, in the standard literature on the equity-efficiency trade-off, as far as the identity of the least-advantaged is known at the time of the contract, the maximin can be only 'imposed' to the most-advantaged; in this sense, we also offer a normative framework by which social contractualism behind the veil of ignorance is formalized in such a way as to ensure the stability of political institutions in a sort of constitutional regime.[25]

As for the identification of optimal social contracts, we show that, due to uncertainty conditions in the matching between preference types and native talents, Rawls-optimal contracts are a subset of Pareto-optimal ones. More specifically, if redistribution is desirable in the concordant-state only, then Rawls optimality implies Pareto optimality, and vice versa. Differently, when redistribution is desirable in both, the concordant and the discordant-state, Rawls optimality implies Pareto optimality, not the converse. Hence, provided that an unanimous agreement on the scheme of wages is required behind the veil of ignorance, i.e. ex-ante, Rawlsian contractualism seems to be a better starting-gate for the refinement of Pareto-optimality, since it introduces distributive justice while preserving unanimity conditions imposed ex-post in the case of Pareto-optimality.

[25] As compared to Phelps (1973), where individuals are assumed to differ from each other with respect to native talent only, in our model individuals also differ in terms of preferences. As such, in our model the possibility of a discordant-state automatically implies uncertainty with respect to the identification of the least-advantaged, whereas this possibility is not conceived in the existing literature. Also, in Phelps (1973), taxation applies to the effective income realized in the labor market which is defined as a function of native talent and effort in education. In our model, instead, native talent and effort in education determine the earnings capacity, i.e. the wage rate, of the individual, which, to the extent that the entire time endowment is regarded as primary good, corresponds to potential, not effective, income.

Finally, even if Rawls' Theory goes beyond distributive justice, it is worth emphasizing major similarities and departures with respect to Roemer's egalitarianism of opportunity (Roemer 1993, 1998). Both approaches are aimed at identifying legitimate inequalities according to some ideal of responsibility. However, within Rawlsian well-ordered society, the *principle of reward*—according to which inequalities due to responsible choices ought not to be compensated—applies to the sole earnings inequalities originating from unequal effort in job (not education). Differently, the *principle of compensation*—according to which inequalities due to circumstances ought to be compensated—never applies in Rawls' framework. Indeed, inequalities cannot originate from social circumstances, i.e. means of resources provided by social institutions (health care, education,...), since these disparities must be eliminated *a priori* according to the principle of Fair Equality of Opportunity. In addition, inequalities originating from natural circumstances, i.e. preference type and native talent, are not to be compensated but regulated according to the maximin principle; "*[t]he intent is not simply to assist those who lose out through accident or misfortune (although that must be done), but rather to put all citizens in a position to manage their own affairs on a footing of a suitable degree of social and economic equality*" (Rawls 2001, p.139).

Appendix

A.1: Proof of Proposition 1

Replace (1) and (7) into (4). Potential wage incomes in the concordant-state of the *HH*- and *LL*-type are, respectively,

$$w_{HH} = -\frac{(\beta - 2)(\theta_H + \theta_L)(a_H(a_L - 1)(\beta - 2)\theta_H - a_H a_L \beta \theta_L + a_L \beta \theta_L)}{16(a_H - 1)(a_L - 1)}$$

$$w_{LL} = \frac{1}{16}(\beta - 2)(\theta_H + \theta_L)\left(\frac{a_H \beta \theta_H}{a_H - 1} - \frac{a_L(\beta - 2)\theta_L}{a_L - 1}\right)$$

where $w_{HH} > w_{LL}$ by construction. Consider the least-advantaged individual.

$$\frac{\partial w_{LL}}{\partial \beta} = 0 \rightarrow \beta^* = \frac{a_H \theta_H (1 - a_L) - 2 a_L \theta_L (1 - a_H)}{a_H \theta_H (1 - a_L) - a_L \theta_L (1 - a_H)}$$

with

$$\frac{\partial^2 w_{LL}}{\partial \beta^2} = \frac{(\theta_H + \theta_L)(a_H(a_L - 1)\theta_H - a_H a_L \theta_L + a_L \theta_L)}{8(a_H - 1)(a_L - 1)} < 0$$

It is easy to verify that $\beta^* \in (0, 1)$ iff $\frac{a_H \theta_H}{1 - a_H} > \frac{2(a_L \theta_L)}{1 - a_L}$, otherwise $\beta^* < 0$. Hence, if $\frac{a_H \theta_H}{1 - a_H} \leq \frac{2(a_L \theta_L)}{1 - a_L}$, then β^* is a negative maximum, which implies that w_{LL} is always decreasing in $\beta \in (0, 1)$, so that $\beta_1^* = 0$. Otherwise, if $\frac{a_H \theta_H}{1 - a_H} > \frac{2(a_L \theta_L)}{1 - a_L}$, then w_{LL} is inverse U-shaped in $\beta \in (0, 1)$ with a single maximum $\beta_1^* = \beta^*$.

Finally, provided that

$$\Delta\Theta_1 = \Theta_{HH} - \Theta_{LL} = \frac{(2-\beta)(\theta_H+\theta_L)}{8(1-a_H)(1-a_L)}[a_H\theta_H(1-a_L) - a_L\theta_L(1-a_H)]$$

the following equivalence condition can be derived

$$\frac{a_H\theta_H}{1-a_H} \gtreqless \frac{2(a_L\theta_L)}{1-a_L} \Leftrightarrow \Delta\Theta_1 \gtreqless \frac{a_L\theta_L(\theta_H+\theta_L)(2-\beta)}{8(1-a_L)} = k_1$$

A.2: Proof of Proposition 2

By replacing (1) and (7) into (4), potential wage income in the discordant-state of the *HL*- and *LH*-type are, respectively,

$$w_{HL} = \frac{1}{16}(\beta-2)(\theta_H+\theta_L)\left(\frac{a_L\beta\theta_H}{a_L-1} - \frac{a_H(\beta-2)\theta_L}{a_H-1}\right)$$
$$w_{LH} = -\frac{(\beta-2)(\theta_H+\theta_L)((a_H-1)a_L(\beta-2)\theta_H - a_Ha_L\beta\theta_L + a_H\beta\theta_L)}{16(a_H-1)(a_L-1)}$$

where, depending on conditions (8), two different cases must be considered; either $w_{HL} > w_{LH}$, or $w_{LH} > w_{HL}$.

Case 1: $w_{HL} > w_{LH}$.

$$\frac{\partial w_{LH}}{\partial\beta} = 0 \to \beta^* = \frac{2(a_H-1)a_L\theta_H - a_Ha_L\theta_L + a_H\theta_L}{(a_H-1)a_L\theta_H - a_Ha_L\theta_L + a_H\theta_L}$$

It is easy to verify that $\beta^* \in (0,1)$ iff $\frac{a_H\theta_L}{1-a_H} > \frac{2(a_L\theta_H)}{1-a_L}$, otherwise $\beta^* < 0$. In addition,

$$\frac{\partial^2 w_{LH}}{\partial\beta^2} = -\frac{(\theta_H+\theta_L)((a_H-1)a_L\theta_H - (a_L-1)a_H\theta_L)}{8(a_H-1)(a_L-1)} < 0$$

This proves that, if $\frac{a_H\theta_L}{1-a_H} \leq \frac{2(a_L\theta_H)}{1-a_L}$, then β^* is a negative maximum, which implies that w_{LH} is always decreasing in $\beta \in (0,1)$, so $\beta^*_{21} = 0$. Otherwise, $\frac{a_H\theta_L}{1-a_H} > \frac{2(a_L\theta_H)}{1-a_L}$, implying that w_{LH} is inverse U-shaped in $\beta \in (0,1)$ with a single maximum $\beta^*_{21} = \beta^*$.

Provided that

$$\Delta\Theta_2 = \Theta_{HL} - \Theta_{LH} = \frac{(2-\beta)(\theta_H+\theta_L)}{8(1-a_H)(1-a_L)}[a_H\theta_L(1-a_L) - a_L\theta_H(1-a_H)]$$

the following equivalence condition is derived

$$\frac{a_H\theta_L}{1-a_H} \gtreqless \frac{2(a_L\theta_H)}{1-a_L} \Leftrightarrow \Delta\Theta_2 \gtreqless \frac{a_L\theta_H(\theta_H+\theta_L)(2-\beta)}{8(1-a_L)} = k_{21}$$

Case 2: $w_{LH} > w_{HL}$.

$$\frac{\partial w_{HL}}{\partial \beta} = 0 \rightarrow \beta^* = \frac{(a_H - 1)a_L\theta_H - 2a_Ha_L\theta_L + 2a_H\theta_L}{(a_H - 1)a_L\theta_H - a_Ha_L\theta_L + a_H\theta_L}$$

It is easy to verify that $\beta^* \in (0, 1)$ iff $\frac{a_L\theta_H}{1-a_H} > \frac{2(a_H\theta_L)}{1-a_L}$, otherwise $\beta^* < 0$. In addition,

$$\frac{\partial^2 w_{HL}}{\partial \beta^2} = \frac{(\theta_H + \theta_L)((1 - a_L)a_H\theta_L - (1 - a_H)a_L\theta_H)}{8(a_H - 1)(a_L - 1)} < 0$$

This proves that, if $\frac{a_L\theta_H}{1-a_L} \leq \frac{2(a_H\theta_L)}{1-a_H}$, then β^* is a negative maximum, which implies that w_{HL} is always decreasing in $\beta \in (0, 1)$, so $\beta^*_{22} = 0$. Otherwise, $\frac{a_L\theta_H}{1-a_H} > \frac{2(a_H\theta_L)}{1-a_L}$, implying that w_{HL} is inverse U-shaped in $\beta \in (0, 1)$ with a single maximum $\beta^*_{22} = \beta^*$.

Provided that

$$(-\Delta\Theta_2) = \Theta_{LH} - \Theta_{HL} = \frac{(2 - \beta)(\theta_H + \theta_L)}{8(1 - a_H)(1 - a_L)}[a_L\theta_H(1 - a_H) - a_H\theta_L(1 - a_L)]$$

the following equivalence condition is derived

$$\frac{a_L\theta_H}{1 - a_L} \gtreqless \frac{2(a_H\theta_L)}{1 - a_H} \Leftrightarrow (-\Delta\Theta_2) \gtreqless \frac{a_H\theta_L(\theta_H + \theta_L)(2 - \beta)}{8(1 - a_H)} = k_{22}$$

A.3: Proof of Corollary 1

Consider first β^*_1 and β^*_{21}. If $\beta^*_1 = 0$, then $\frac{a_H\theta_H}{1-a_H} \leq \frac{2a_L\theta_L}{1-a_L}$, which automatically implies $\frac{a_H\theta_L}{1-a_H} \leq \frac{2a_L\theta_H}{1-a_L}$, so $\beta^*_{21} = 0$. Suppose $\frac{a_H\theta_H}{1-a_H} = \frac{2a_L\theta_L}{1-a_L}$, then there exists $\epsilon > 0$ small enough such that $\frac{a_H\theta_H}{1-a_H} + \epsilon > \frac{2a_L\theta_L}{1-a_L}$, implying $\beta^*_1 > 0$ and $\beta^*_{21} = 0$. Let $\beta^*_{21} > 0$ be the optimal state-contingent contract when w_{LH} is inverse U-shaped in $\beta \in (0, 1)$. From conditions above, $\beta^*_{21} > 0$ implies $\beta^*_1 > 0$ and, mostly,

$$\beta^*_1 - \beta^*_{21} = \frac{a_Ha_L(1 - a_H)(1 - a_L)(\theta_H^2 - \theta_L^2)}{\big(a_L\big(1 - a_H\big)(\theta_H - \theta_L)\big)\big(a_H\theta_H(1 - a_L) - a_L\theta_L(1 - a_H)\big)} > 0$$

Consider now β^*_1 and β^*_{22}. For the same argument above, $\beta^*_{22} > 0$ implies $\beta^*_1 > 0$, whereas the opposite is not true. In addition, let β^*_{22} be the optimal state-contingent contract when w_{HL} is inverse U-shaped in $\beta \in (0, 1)$, then

$$\beta^*_1 - \beta^*_{22} = \frac{\theta_H\theta_L\big((a_H - a_L)(a_H + a_L - 2a_Ha_L)\big)}{((a_H - 1)a_L\theta_H - a_Ha_L\theta_L + a_H\theta_L)(a_H(a_L - 1)\theta_H - a_Ha_L\theta_L + a_L\theta_L)}$$

which is positive when $a_H(1 - 2a_L) > -a_L$. Clearly, this condition is always satisfied for all $a_L \in (0, 1/2)$. In addition, for all $a_L \in [1/2, 1)$, it holds if $a_H < \frac{a_L}{2a_L-1}$, where $a_L \geq 2a_L - 1 \;\forall\; a_L \leq 1$ implies that $a_H \leq 1 \leq \frac{a_L}{(2a_L-1)}$. This proves that $a_H(1 - 2a_L) > -a_L$ is always satisfied $\forall\; a_L < 1$, so that $\beta^*_1 > \beta^*_{22}$.

A.4: State-contingent Rawls-efficiency frontier

In what follows, we construct the Rawls-efficiency frontier for the concordant-state. For the sake of brevity, the same procedure is omitted for the discordant-state, but it is available upon request.

Recall w_{HH} from Appendix A.1 and consider the first-order condition in the concordant-state,

$$\frac{\partial w_{HH}}{\partial \beta} = 0 \rightarrow \beta^* = \frac{2a_H\theta_H(1-a_L) - a_L\theta_L(1-a_H)}{a_H\theta_H(1-a_L) - a_L\theta_L(1-a_H)} > 1$$

Since

$$\frac{\partial^2 w_{HH}}{\partial \beta^2} = \frac{(\theta_H + \theta_L)(a_H\theta_H(1-a_L) - a_L\theta_L(1-a_H)}{8(1-a_L)(1-a_H)} > 0$$

it must be the case that $\beta^* > 1$ is a minimum, implying that w_{HH} is monotonically decreasing in $\beta \in (0,1)$.[26]

Hence, take the inverse function $\beta(w_{HH})$ and replace it in w_{LL} from Appendix A.1. The Rawls-efficiency frontier for the concordant-state is then

$$\begin{aligned} w_{LL} = & \frac{\left(a_L\left(\Gamma[\cdot] + \theta_H\theta_L + \theta_L^2\right) - \Gamma[\cdot]\right)}{16(1-a_L)(\theta_H+\theta_L)(a_H\theta_H(1-a_L) - a_L\theta_L(1-a_H))} \\ & \left(a_H\left(a_L\Gamma[\cdot]\ -\Gamma[\cdot] + 2(a_L-1)\theta_H^2 + (3a_L-2)\theta_H\theta_L + a_L\theta_L^2\right) + \right. \\ & \left. -a_L\Gamma[\cdot] + \Gamma[\cdot] - a_L\theta_H\theta_L - a_L\theta_L^2\right) \end{aligned}$$

where

$$\begin{aligned} & \Gamma[a_H, a_L, \theta_H, \theta_L, w_{HH}] = \\ & \sqrt{\frac{(\theta_H+\theta_L)\left((a_H-1)(\theta_H+\theta_L)a_L^2\theta_L^2 - 16w_{HH}(a_L-1)\left(a_L\theta_L(1-a_H) + a_H\theta_H(a_L-1)\right)\right)}{(a_H-1)(a_L-1)^2}} \end{aligned}$$

Funding Open access funding provided by Università degli Studi di Salerno within the CRUI-CARE Agreement.

[26] It is easy to verify (and intuitive) that the same monotonicity condition holds true in the two discordant-states as well.

References

Alexander, S. S. (1974). Social evaluation through notional choice. *The Quarterly Journal of Economics, 88*(4), 597–624.

Amiel, Y., & Cowell, F. (1994). Monotonicity, dominance and the Pareto principle. *Economics Letters, 45*, 447–450.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy, 118*(4), 649–680.

Hammond, P. J. (1981). Ex-ante and ex-post welfare optimality under uncertainty. *Economica, 48*(191), 235–250.

Hampton, J. (1980). Contracts and choices: Does rawls have a social contract theory? *Journal of Philosophy, 77*(6), 315–338.

Harris, R. G. (1978). Ex-post efficiency and resource allocation under uncertainty. *Review of Economic Studies, 45*, 427–436.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy, 61,* 434–435.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy, 63*, 309–321.

Holmstrom, B. (1982). Moral Hazard in Teams. *The Bell Journal of Economics, 13*(2), 324–340.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies, 38*(2), 175–208.

Moehler, M. (2015). The Rawls–Harsanyi dispute: A moral point of view. *Pacific Philosophical Quarterly, 99,* 82–99.

Mongin, P. (2001). The impartial observer theorem of social ethics. *Economics and Phylosophy, 17*, 147–179.

Musgrave, R. A. (1974). Maximin, uncertainty, and the leisure trade-off. *The Quarterly Journal of Economics, 88*(4), 625–632.

Phelps, E. S. (1973). Taxation of wage income for economic justice. *The Quarterly Journal of Economics, 87*(3), 331–3554.

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

Rawls, J. (1974). Reply to Alexander and Musgrave. *The Quarterly Journal of Economics, 88*(4), 633–655.

Rawls, J. (1975). Fairness to goodness. *The Philosophical Review, 84*(4), 536–554.

Rawls, J. (2001). *Justice as fairness: A restatement*. E. Kelly (ed.), Cambridge, MA: Harvard University Press.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy and Public Affairs, 22*(2), 146–166.

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

Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature, 54*(4), 1288–1332.

Rousseau, J.J. *Du contrat social: ou principes du droit politique*. Collection complète des oeuvres (G.D.H. Cole translation), Genève, 1762.

Saposnik, R. (1981). Rank-dominance in income distributions. *Public Choice, 36*(1), 147–151.

Sen, A. (1992). *Inequality Reexamined*. Oxford: Clarendon Press.

Starr, R. (1973). Optimal production and allocation under uncertainty. *The Quarterly Journal of Economics, 87*, 81–95.

Stiglitz, J.E. (1987). *Pareto efficient and optimal taxation and the new new welfare economics*. Ch. 15 in: Handbook of Public Economics vol. 2, **87**, 991–1042.

Sugden, R. (1993). Welfare, resources and capabilities: A review of inequality reexamined by Amartya Sen. *Joural of Economic Literature, 31,* 1947–1962.

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