ECINEQ
Society for the Study of Economic Inequality

Working Paper Series

The economics of justice as fairness

Antonio Abatemarco
Francesca Stroffolini

ECINEQ WP 2017 - 430

ECINEQ 2017 - 430
March 2017

www.ecineq.org

The economics of justice as fairness

Antonio Abatemarco†
University of Salerno, Italy

Francesca Stroffolini
University of Naples "Federico II", Italy

Abstract

In this paper we challenge the common interpretation of Rawls' Theory of Justice as Fairness by showing that this Theory, as outlined in the Restatement (Rawls 2001), goes well beyond the definition of a distributive value judgment, in such a way as to embrace efficiency issues as well. A simple model is discussed to support our interpretation of the Difference Principle, by which inequalities are shown to be permitted as far as they stimulate a greater effort in education in the population, and so economic growth. To our knowledge, this is the only possibility for the inequality to be 'bought' by both the most-, and above all, the least-advantaged individual as suggested by the Difference Principle. Finally, by recalling the old tradition of universal ex-post efficiency (Hammond 1981), we show that a unique optimal social contract does not exist behind the veil of ignorance; more precisely, the sole set of potentially optimal social contracts can be identified a priori, and partial justice orderings derived accordingly.

†**Contact details:** A. Abatemrco (corresponding author), Department of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132 - 84084 - Fisciano (SA), Italy; Email: aabatemarco@unisa.it. F. Stroffolini, Department of Economics and Statistics, University of Naples "Federico II", Complesso Universitario di Monte Sant'Angelo, Via Cintia, 21 - 80126 - Napoli, Italy; Email: stroffol@unina.it.

1 Introduction

Starting from Kaldor's (1957) seminal paper, a huge research effort has been devoted to the analysis of the impact of inequality on growth. While the neoclassical theory of economic growth was strictly concerned with the effect of income disparities on physical capital, the recent literature has been mostly focusing on the role of human capital as one of the key factors for a better understanding of inequality implications on growth, especially in advanced economies (e.g., Galor and Zeira 1993, Galor and Moav 2004). Mostly, it is said, in the presence of credit market constraints, inequality may jeopardize access to (pareto-efficient) investments in human capital and, in turn, obstruct economic growth.

In this paper we propose a reinterpretation of Rawls' Theory of Justice as Fairness (hereafter Theory), by which the causality between inequality, individual investments in human capital, and economic growth is modeled 'behind the veil of ignorance'. Specifically, within our reinterpretation of the Theory, distributive aspects are assumed to impact on growth to the extent that the magnitude of (expected) income inequalities is said to influence (ongoing) individual incentives to effort in education, and so (future) productivity in the labor market.

According to the common understanding of the Theory, an allocation is to be preferred if and only if the 'least-advantaged' individual is better off, independently from efficiency issues; this is the main idea usually ascribed to the maximin principle as represented by the well known Leontief preferences (Alexander 1974). In our view, however, such interpretation would not leave any room for the Rawlsian Difference Principle, by which, it is said, once education opportunities are granted to the entire population, inequalities are admitted as far as they are to the greatest benefit of the least-advantaged. Evidently, in the absence of efficiency issues, there is no way by which the least-advantaged individual might be willing to be penalized by the introduction of inequality. As such, for any inequality to represent a benefit for the worse-off (i.e. least-advantaged), economic growth must be necessarily accounted for, that is, inequality must be stimulating growth in such a way as to make the least-advantaged, for some degree of inequality at least, more than compensated for being the worse-off. In a sense, inequality must be aimed at pro-poor growth.

In this way of thinking — as stated by Rawls' in the premise of the Restatement published thirty years after the Theory (Rawls 2001)[1] — a revision of the common understanding of the Theory is necessary because, in our view, this Theory goes well beyond the proposal of a distributive value judgment in such a way as to embrace efficiency issues as well[2].

[1] "In this work I have two aims. One is to 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 ..." (Rawls 2001, p.xv).

[2] "[I]t is not correct, I think, 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).

According to our economic reinterpretation of Rawls' thought, inequalities influence individual incentives to effort in education, and so the earnings ability they will realize in the labor market. Most importantly, the earnings ability is said to be determined as a result of both effort in education and native abilities, so that the most-advantaged individual does not necessarily correspond to the better endowed in terms of native abilities.

To better support our reinterpretation of the Theory, we model the Rawlsian Difference Principle through a two stages sequential game, where knowledge available to the parties is progressively relaxed over time. Precisely, we focus on the sole Difference Principle by assuming that the Rawlsian principle of Fair Equality of Opportunity holds, so that access to education and training is assumed to be universally granted independently from the social class of origin.

Within this framework, we assume that at time 0, i.e. behind the veil of ignorance, the two (groups of) individuals define the scheme of wages and salaries, that is, the wage rate to be paid in the labor market as a function of the earnings ability realized until the age of reason. Behind the veil of ignorance, individuals have no information about both (i) their preferences (i.e., propensity to effort in education) and (ii) their native talent. At time 1, given the scheme of wages and salaries agreed at time 0, the preferences of both individuals reveal (complete information). At this stage, individuals choose their effort in education as conditioned to the propensity to effort of the other individual (Nash bargaining).

Solving by backward induction, we show that a unique optimal social contract cannot exist behind the veil of ignorance; as far as the individual with the higher propensity to effort in education might be associated, ex-post, to the better or the worse endowment in terms of native talent, two different states — with two different optimal schemes of wages and salaries — are obtained. Nevertheless, even if a unique optimal social contract cannot be identified behind the veil of ignorance, we show that the set of potentially optimal social contracts can be still determined in the perspective of universal ex-post efficiency (Starr 1973, Harris 1978, Hammond 1981), and partial 'justice' orderings derived accordingly.

The paper is organized as follows. In Section 2 we present our reinterpretation of the Theory by recalling the definition of the original position and the two Rawlsian Principles. The model formalizing our interpretation of the Theory is discussed in Section 3. Accordingly, the set of optimal social contracts is derived in Section 4. Section 5 concludes.

2 The Theory of Justice

2.1 Contractualism in the original position

In His Theory (1971), John Bordley Rawls proposes a political conception of justice by which the stability of political institutions is to be preserved by ensuring the *overlapping consensus* in the society; remarkably, the overlapping consensus is said to grant the stability of the society, independently of the oppressive sanctions of state power which, instead, are inevitably required in a society united on a form of utilitarianism.[3]

In this sense, the Theory of Justice of Fairness is usually accommodated in the old tradition of *social contractualism* whose best known proponents are Hobbes, Locke and Rousseau. Specifically, Rawls explores the possibility of a social contract to be agreed in the original position, or, equivalently, behind the veil of ignorance.

The focus on the original position is crucial 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.[4]

More precisely, three conditions — fundamental for our interpretation — 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 their individualistic preferences, but these are taken as 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 is not given in the original position. Most importantly, to the extent that 'techniques at disposal of the society' are unknown at this stage, native talent is merely potential and not measurable apart from social institutions revealing ex-post; e.g., the same native endowment may be more or less successful in the society depending on social and other contingencies.[5] Third (c), the social contract is agreed under uncertainty conditions where the lack of information is radical, so 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 all this information (i.e., a-b-c), it must be the case that, in the original position, no one is advantaged or disadvantaged by natural chance or social contingencies

[3] "A society united on a form of utilitarianism, or on the moral views of Kant or Mill, would likewise require the oppressive sanctions of state power to remain so" (Rawls 2001, p.34). On the conflict between Rawlsian theory and the philosophical underpinnings of utilitarianism see also Bradford (2012).

[4] "[S]tability results first, from the availability of principles that guarantee citizens' fundamental interests" (Rawls 2001, p.110).

[5] "[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.

in the choice of principles, which is a *conditio sine qua non* for the overlapping consensus to hold. Notably, at this stage, individuals are supposed to decide the principles they are willing to adopt, and not the most effective means to one's ends.[6]

What is known behind the veil of ignorance, instead, is the object of 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 perspective, for instance, income and wealth are said to belong to the set of primary goods to the extent that they are fundamental to implement a political conception of the person as free and equal, endowed with the moral powers, and capable of being a fully cooperating member of the society.[7]

Within the Restatement (Rawls 2001), most of the emphasis is posed on the notion of 'lifetime income prospect', which is intended as the synthetic measure, or index, quantifying the primary goods an individual may have access to over a complete life, e.g. the income an individual can reasonably expect (e.g., prospect) when the age of reason is achieved. Differences in citizens' lifetime income prospects are said to be influenced by such things as their social class of origin, their native endowments, their ambitions, their opportunities for education, and their good or ill fortune over the course of life.

Most importantly, individuals are expected to receive a share of lifetime income prospect depending on their training/education effort, that is, on how much they have contributed "*by training and educating their native endowments putting them to work within a fair system of social cooperation*" (Rawls 2001, p.68). In this sense, the 'realized talent' — which is measurable as opposed to the 'native talent' that is not — is intended as the result we may have after we have reached the age of reason, and it is inevitably determining the amount of primary goods an individual would have access to (over a complete life) in the absence of any scheme of social cooperation, that is, when redistribution is not allowed in any form.

Diversely, when a scheme of social cooperation is agreed behind the veil of ignorance, the lifetime income prospect of each individual is not necessarily anchored to its own realized talent any longer; in a cooperative system, the

[6] "We suppose that the parties are rational, where rationality (as distinguished from reasonableness) is understood in the way familiar from economics. Thus the parties are rational in that they can rank their final ends consistently; they deliberate guided by such principles as: to adopt the most effective means to one's ends; to select the alternative most likely to advance those ends; to schedule activities so that, ceteris paribus, more rather than less of those ends can be fulfilled" (Rawls 2001, p.87).

[7] "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).

scheme of wages and salaries is expected to allocate among its own members the overall amount of talent realized (e.g. productivity, earnings ability) in the society as a whole. In this scenario, the lifetime income prospect of each individual is inevitably determined by a mix of its own realized talent and the one realized by others, in a way that embodies some redistribution from the most to the least-advantaged (as identified in terms of lifetime income).

2.2 The two principles of justice

Given the very basic set up characterizing the original position, Rawls suggests two principles which, in His view, would make differences in lifetime income prospects 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.

By the former, it is said, "[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 latter, "[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, because it automatically implies that equality cannot be pursued through progressive taxation as this would violate the Liberty principle.[8] More precisely, the redistribution of wealth and income can be admitted exclusively to prevent excessive concentrations of property and wealth, especially those likely to lead to political domination, as they would threaten the political liberties, i.e. the basic liberties safeguarded by the first principle. Indeed, even if the social contract is defined according to the two principles of justice, excessive concentrations may still come out from bequests and inheritance, as well as from separate and seemingly fair agreements between individuals, which would inevitably jeopardize the overlapping consensus.[9]

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; more specifically, once political institutions have granted similar education and

[8] "For our purposes, however, the relevant difficulty is that a head tax would violate the priority of liberty. It would force the more able into those occupations in which earnings were high enough for them to pay off the tax in the required period of time; it would interfere with their liberty to conduct their life within the scope of the principles of justice" (Rawls 2001, p.158).

[9] "[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).

training possibilities for all members of the society, the social contract is to be designed in such a way as to permit the sole inequalities benefitting the least-advantaged. Notably, the principle of Equality refers to the sole inequalities defined in terms lifetime income prospect in that, by virtue of *background procedural justice*, the ex-post distribution of income and wealth — as obtained according to the two principles — is to be regarded just (or at least not unjust) whatever this distribution turns out to be.

To the extent that, by virtue of the principle of Fair Equality of Opportunity, political institutions are supposed to neutralize the impact of the social class of origin on the education and training opportunities (given that good or ill fortune is normally distributed), by recalling the Rawlsian definition of lifetime income prospect from the previous Section, it must be the case that the only disparities of lifetime income prospects (yet not legitimate) must be originating from different talent and/or propensity to effort in education (i.e., preferences).

As a consequence — and to our knowledge this aspect has not been properly emphasized in the common understanding of Rawls' thought — worse endowments in terms of native talent do not necessarily imply lower lifetime income prospects (i.e.. least-advantaged), because native endowments must first be realized through educational and training effort, which belongs to the private sphere of individual decisions.[10] In addition, to the extent that individual decisions matter, the social contract — as defined behind the veil of ignorance — is not to be intended as merely redistributive but, also, the mechanism-design by which incentives to effort in education are defined.

In what follows, we model our interpretation of the Theory in such a way as to account for the implications of inequalities on individual decisions of effort in education, and so economic growth.[11] According to the Difference Principle, inequalities are legitimate to the extent that they induce growth which is benefitting the least-advantaged; to the extent that some inequalities may induce growth that is penalizing the least-advantaged, not all 'growth-enhancing' inequalities are admissible. This poses a precise limit on the maximum inequality admissible in the society which, in a way, evokes the ideal of pro-poor growth.

3 The Model

In this section we discuss a simple analytical framework by which the Rawlsian theory - as revisited above - is formalized. More precisely, the Difference Principle is modeled once, according to the priority assigned by Rawls within the Equality principle, conditions of Fair Equality of Opportunity are taken for

[10] "[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).

[11] "Thus the principles of social justice are macro and not necessarily micro principles" (Rawls 1974).

granted for the entire population, meaning that, education/training opportunities are assumed to have already been equalized for all individuals, so that the social class of origin can be omitted.

Given a population of two individuals, let $\theta^i, \theta^j \in \Re^+$ be the native talent of the ith and jth individual respectively. According to Rawls theory, native endowments can be inferred ex-post only, in that this value is not measurable in itself ex-ante, and highly dependent on the design of social institutions revealing ex-post. As such, given a population of two individuals who are assumed to differ in terms of native talent, with $\theta_H > \theta_L$, two different states of the world are to be considered. Depending on social institutions, either $\theta^i = \theta_H$ and $\theta^j = \theta_L$, or $\theta^i = \theta_L$ and $\theta^j = \theta_H$. Most importantly, as we observed in the previous Section, behind the veil of ignorance the probability is intended in classical terms, so that the two states are equally probable.

Given the native talent, let Θ be the realized talent 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 indicating the effort in education. For our purposes, Θ is assumed to indicate the money-value of the realized talent, which might be though as a sort of individual productivity determined by education decisions and native talent.

Let ℓ be the individual lifetime income prospect indicating the primary goods an individual may potentially have access to. Most importantly, an aspect we will have to come back later on, the lifetime income prospect is a potential value which is defined up to the entire time endowment, leisure included, of each individual.[12] Formally, $\ell^i = \Theta^i T$ and $\ell^j = \Theta^j T$ with T indicating the time endowment.

In the absence of redistribution, the lifetime income prospect is fully determined by the realized talent, so that, assuming $T = 1$ without loss of generality, $\ell^i = \Theta^i$ and $\ell^j = \Theta^j$. Differently, when redistribution is allowed behind the veil of ignorance, given the budget constraint $\ell^i + \ell^j = \Theta^i + \Theta^j$, the lifetime income prospect of each individual is not anchored any longer to the corresponding realized talent.

In what follows, for the sake of simplicity, we focus on the case of a linear redistributive system, so that

$$\begin{aligned} \ell^i &= \alpha + (1 - \beta)\Theta^i \\ \ell^j &= \alpha + (1 - \beta)\Theta^j \end{aligned} \tag{2}$$

[12] "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).

where $\alpha > 0$ and $\beta \in [0, 1]$ identifies the scheme of wages and salaries (or, equivalently, scheme of social cooperation). Remarkably, to the extent that the budget constraint is required to hold, i.e. $\ell^i + \ell^j = \Theta^i + \Theta^j$, from (2) it must be the case that $\alpha = (\beta/2)(\Theta^i + \Theta^j)$, so that (2) can be equivalently rewritten as

$$\begin{aligned} \ell^i &= \frac{\beta}{2}\Theta^j + \left(1 - \frac{\beta}{2}\right)\Theta^i \\ \ell^j &= \frac{\beta}{2}\Theta^i + \left(1 - \frac{\beta}{2}\right)\Theta^j \end{aligned} \tag{3}$$

so that, the higher is β, the greater is the contribution to the ith lifetime income prospect of the jth realized talent, and vice versa.

From (2) and (3), any increase of β is both (i) increasing the redistribution of realized talents across the population, i.e. reducing inequality in the distribution of lifetime income prospects, and (ii) reducing the return to effort, i.e. lowering the incentive to effort in education. Also, it is worth observing that the redistribution originating from the scheme of wages and salaries is ordering-preserving by construction, in that the identification of the least-advantaged individual, i.e. the sign of $(\ell^i - \ell^j) = (\theta^i - \theta^j)$, is independent of β.

Once the realized talent and the lifetime income prospect are defined, let's turn to the timing of the game. At time 0, the two (groups of) individuals define the scheme of wages and salaries behind the veil of ignorance, that is without any information about their native talent and preferences. At time 1, individual preferences reveal for both individuals; at this stage, each individual is supposed to choose effort in education in such a way as to maximize the expected lifetime income (i.e., lifetime income prospect), given the scheme of wages and salaries defined at the previous stage.

Given a population of two individuals who are assumed to differ in terms of the mutually-exclusive native talents, $\theta_H > \theta_L$, two different states of the world are to be considered at time 1; either the ith individual is associated to θ_H whereas the jth individual is of the type θ_L, or vice versa. Most importantly, at time 1, the probability is intended in classical terms, so that, like in the case of preferences at time 0, the two states of the world are taken as equally probable.

Thus, the optimal scheme of wages and salaries 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.

3.1 Utility maximization

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 utility.[13] Specifically, we define the *linear*[14] utility as a function of the

[13]Precisely, even if the concept of utility is specifically mentioned when offering the interpretation of the Difference Principle (Rawls 2001,p.), to our knowledge, there is no explicit reference to the standard utility maximization framework.

[14]As far as the utility is defined with respect to the expected lifetime income, risk neutrality is implicitly assumed, independently of the linearity of the utility function. In our view, as it

share[15] of the expected lifetime income, i.e. lifetime income prospect, and the dis-utility from effort in education.

Let $a^i, a^j \in [0,1]$ indicate the relative contribution (i.e. marginal utility) of the share of lifetime income prospect to the overall utility of the ith and the jth individual respectively. We assume that individuals differ from each other in terms of the mutually-exclusive propensities to effort in education. Thus, given $a_H > a_L$, once the ith preferences have revealed, it must be the case that the jth preferences can be inferred by the ith individual, and vice versa (i.e., complete information). To simplify the formalization, we hypothesize $a^i = a_H$ (so, $a^j = a_L$), as the opposite case implies perfectly symmetric definitions. As a result, the utilities of the two individuals are defined as follows

$$\begin{aligned} U^i &= a^i \left(\tfrac{E[\ell^i]}{E[\ell^i]+E[\ell^j]} \right) + (1-a^i)(1-e^i) \\ U^j &= a^j \left(\tfrac{E[\ell^j]}{E[\ell^i]+E[\ell^j]} \right) + (1-a^j)(1-e^j) \end{aligned} \tag{4}$$

where $(1-e)$ is the dis-utility from effort, which is assumed to be linear for the sake of simplicity, and $E[\ell]$ is the expected lifetime income. Specifically, to the extent that possible states are taken as equally probable within the Rawlsian Theory, the lifetime income prospects of the two individuals are defined with respect to the (mutually-exclusive) native talents revealing ex-post, i.e. $\theta_H > \theta_L$, as follows,

$$\begin{aligned} E[\ell^i] &= \tfrac{1}{2}\ell^i_H + \tfrac{1}{2}\ell^i_L \\ E[\ell^j] &= \tfrac{1}{2}\ell^j_L + \tfrac{1}{2}\ell^j_H \end{aligned} \tag{5}$$

where ℓ^i_k, ℓ^j_k, $k = H, L$, are, respectively, the ith and the jth state-contingent lifetime income prospect, as obtained by replacing (1) in (3) with $\theta^i, \theta^j = \theta_k$, $k = H, L$ and $\theta^i \neq \theta^j$.

Therefore, as far as (i) complete information is assumed on preferences and (ii) individuals are supposed to act simultaneously, Nash-equilibria are considered. More precisely, from the utility maximization the two reaction functions are defined as follows,[16]

$$\begin{aligned} e^{i*} &= -e^j + \sqrt{\frac{a_H}{1-a_H} e^j (1-\beta)} \\ e^{j*} &= -e^i + \sqrt{\frac{a_L}{1-a_L} e^i (1-\beta)} \end{aligned} \tag{6}$$

whose main characteristics are outlined in Property 3.1.

will be clearer later on, risk aversion is far from essential in Rawls' Theory; "[i]t is tempting at first sight to suppose that the maximin criterion is based on an extreme and arbitrary assumption about risk aversion. I wish to show that this is a misapprehension".

[15] "The two principles of justice assess the basic structure according to how it regulates citizens' shares of primary goods, these shares being specified in terms of an appropriate index (Rawls 2001, p.59)".

[16] As far as effort is non-negatively defined, the sole positively defined reaction functions are considered.

Property 3.1. *Both reaction functions are $\cap-$shaped with maximum value at, respectively, $Max\{e^{i*}\} = \frac{a_H(1-\beta)}{4(1-a_H)}$ and $Max\{e^{j*}\} = \frac{a_L(1-\beta)}{4(1-a_L)}$.*

Proof. See Appendix A.1. □

Intuitively, if the effort of the jth individual increases, then the corresponding increase of its own share of lifetime income prospect is larger when j holds the greater lifetime income prospect ($\ell^j > \ell^i$), or, equivalently, the higher effort ($e^j > e^i$), as compared to the ith individual (see Appendix A.1). As such, the jth individual has an incentive to react by increasing effort as far as its reaction function is above the bisectrix line (i.e., $e^j > e^i$, or, equivalently, $\ell^j > \ell^i$).

By comparing the two maximum values of the two reaction functions, it can be observed that the ith individual, who's the one with a higher propensity to effort ($a_H > a_L$), maximizes its output at a greater effort level as compared to the jth individual (i.e., $e^{i*} > e^{j*}$), even if, according to (3), such a gap is decreasing with redistribution (i.e., when β increases).

From the two reaction functions in (6), two different Nash-equilibria are obtained, respectively,

$$e^*_H = 0; \quad e^*_L = 0$$
$$e^*_H = \frac{a_H^2(1-a_L)a_L(1-\beta)}{(a_H + a_L - 2a_H a_L)^2}; \quad e^*_L = \frac{(1-a_H)a_H a_L^2(1-\beta)}{(a_H + a_L - 2a_H a_L)^2} \tag{7}$$

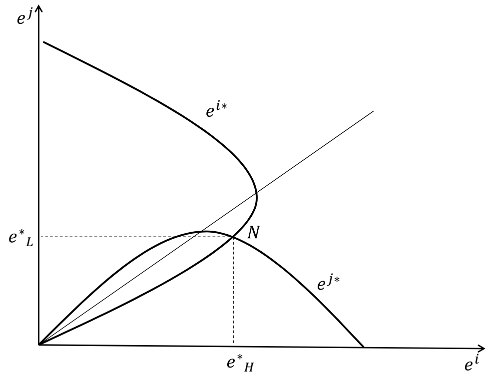
Proposition 3.2. *From the utility maximization, two different Nash equilibria are obtained: the first is characterized by levels of effort equal to zero (so, zero lifetime income prospects); the second, with positive levels of effort (so, positive lifetime income prospects), occurs on the increasing line of the reaction function of the individual with a higher propensity to effort and on the decreasing line of the reaction function of the other individual.*

Proof. See Appendix A.2. □

From Proposition 3.2, it is worth observing that the (interior) Nash equilibrium must occur below the bisectrix (see Fig. 1), meaning that the individual with a better propensity to effort, and thereby who exerts higher effort, must be the one with a larger share of expected lifetime income prospect. This is obvious for all possible schemes of wages and salaries (β), because at this stage both individuals acts on the same levels of expected abilities.

In what follows, we focus on the sole interior Nash equilibrium (N in Fig. 1) in that, according to Rawls (2001), "[o]*ther things being equal, the difference principle directs society to aim at the highest* [lifetime income prospect for the least-advantaged from the] *most effectively designed scheme of cooperation.*" Indeed, as shown in the proof of Proposition 3.2, the interior Nash equilibrium is clearly dominating the corner one for all $a_H, a_L, \beta \in [0, 1]$.

Figure 1: Reaction functions



3.2 State-contingent optimal social contracts

Given the optimal effort individuals are willing to exert, the optimal scheme of wages and salaries can be defined by solving backward, i.e., behind the veil of ignorance.

Let the ith individual be the one endowed with the higher propensity to effort. To the extent that the two individuals are assumed to 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 ith 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 ith individual reveals of type θ_L (implying j's θ_H-type) which we refer to as 'discordant-state'. As such, in the concordant-state the ith individual is the "most-advantaged", whereas the other individual is the "least-advantaged". Differently, in the discordant-state, the least-advantaged cannot be identified *a priori* as 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 realized ability of the ith (resp. jth) individual with higher (resp. lower) propensity to effort and better (resp. worse) native talent, as obtained by replacing the interior Nash equilibrium in (1) with $e^i = e_H$, $e^j = e_L$, $\theta^i = \theta_H$, $\theta^j = \theta_L$. Clearly, $\Theta_{HH} > \Theta_{LL}$.

In the discordant-state, let Θ_{HL} (resp. Θ_{LH}) be the realized ability of the ith (resp. jth) individual with higher (resp. lower) propensity to effort and worse (resp. better) native talent, as obtained by replacing the interior Nash equilibrium in (1) with $e^i = e_H$, $e^j = e_L$, $\theta^i = \theta_L$, $\theta^j = \theta_H$. The least-advantaged individual may be endowed with the lower propensity to effort but better native talent, i.e. $\Theta_{HL} > \Theta_{LH}$. Alternatively, the least-advantaged individual may be the one endowed with higher propensity to effort but worse native talent, i.e. $\Theta_{HL} < \Theta_{LH}$. Formally, recall that (i) the least-advantaged is defined as the individual with the lowest lifetime income prospect, and (ii) the identification of the least-advantaged is independent of β in that (3) is ordering-preserving by construction. As a result, it must be the case that $\Theta_{HL} \gtreqless \Theta_{LH}$ implies $\ell_{HL} \gtreqless \ell_{LH}$, and vice versa. Therefore, it can be shown that, in the discordant-state, the least-advantaged individual is identified by the following (equivalence) condition.

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

To determine the optimal social contract, let's emphasize that any variation of the scheme of wages and salaries (β) generates two different effects on the two lifetime income prospects. On the one hand, according to (3), any increase of β implies a redistribution in terms of realized talent from the most to the least-advantaged type, meaning that β is a redistributive parameter (direct effect). On the other hand, $(1-\beta)$ acts as a sort of wage-premium determining the incentive to effort, and so the lifetime income prospect of both individuals (indirect effect); if β increases, then the relative contribution of the ith (jth) realized talent to its lifetime income decreases, so that any individual is less willing to make high effort in education (Appendix A.3). In this sense, a dis-incentive effect is to be considered too.

Proposition 3.3. *If β increases, then the lifetime income prospect of the most-advantaged individual decreases for all $\beta \in [0,1]$, whereas the lifetime income prospect of the least-advantaged is either (i) decreasing, if the dis-incentive effect dominates the redistributive effect for all $\beta \in [0,1]$, or (ii) $\cap$-shaped, if the redistributive effect dominates the dis-incentive effect for β sufficiently low.*

Proof. See Appendix A.3. □

Proposition 3.3 holds for both the concordant and the discordant-state even if, as reported in Appendix A.4, the two dominance conditions differ from each other in magnitude. Evidently, whatever the state, the lifetime income prospect of the most-advantaged individual is decreasing in β in that, for this individual, the dis-incentive and the redistributive effect are equally-signed (negative). Differently, these two effects are contrasting to each other for the least-advantaged individual. For the latter individual, it turns out that the lifetime income prospect is decreasing when the gap in terms of realized talent (as determined by preferences and native talent) is not large enough, so that the redistributive effect is smaller. On the contrary, when the gap of realized talent is large enough (e.g., $a_H\theta_H > 3a_L\theta_L$ is a sufficient condition), the redistributive effect is dominating, so that the lifetime income prospect of the least-advantaged is initially increasing in β. At this stage, any marginal increase of β reduces the gap between the two lifetime income prospects, so that the redistributive effect becomes weaker and weaker, until the dis-incentive effect becomes dominating. From this point on, both lifetime income prospects decrease in β.

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 lifetime income prospects; i.e., for the least-advantaged, the effect of the minor redistribution is initially more than compensated by the increasing incentive to effort. As such, a marginal decrease of β from $\beta = 1$ generates pareto-improvements (and so, economic growth intended as an increase in the total lifetime income prospect) which are bought by both individuals. Subsequently, once the break-even point is achieved, for any additional increase of inequality, the effect of the minor redistribution becomes dominating for the least-advantaged individual, so that its lifetime income prospect decreases. From now on, any additional increase of inequality — even if it might be growth enhancing — is not bought by the least-advantaged individual in that, growth is not of the pro-poor kind.

To the extent that the social contract is to be bought by both individuals, the optimal social contract (β^*) is obtained when the lifetime income prospect of the least-advantaged individual is maximum. Evidently, this solution is expected to differ depending on the state of the world revealing ex-post. Formally, let $\beta_1^*, \beta_{21}^*, \beta_{22}^* \in [0, 1]$ be the optimal schemes of wages and salaries, respectively, in the concordant-state and in the discordant-case, with β_{21}^* holding if (in the discordant-state) the least-advantaged individual is the better endowed in terms of native talent, and β_{22}^* holding if (in the discordant-state) the least-advantaged individual is the worse endowed in terms of native talent. From Proposition 3.3, two different alternative schemes are to be considered, one of the two being the interior solution depending on the state of the world.

Proposition 3.4. *If the lifetime income prospect of the least-advantaged individual is strictly decreasing in β, then $\beta_1^*, \beta_{21}^*, \beta_{22}^* = 0$. On the other hand, if the lifetime income prospect of the least-advantaged individual is $\cap$-shaped with respect to β, then $\beta_1^* > \beta_{21}^* > \beta_{22}^* > 0$.*

Proof. See Appendix A.4. □

As such, for redistribution to be desirable behind the veil of ignorance (i.e., $\beta > 0$), the lifetime income prospect of the least-advantaged must be $\cap$-shaped. In addition, as far as conditions for the $\cap$-shaped pattern are found to be more stringent in the discordant-state (and within the latter more stringent when $\Theta_{HL} < \Theta_{LH}$), it must be the case that the greater is inequality of realized talent originating from endowments (preferences and native abilities), (i) the more desirable redistributive plans are likely to be (i.e., $\beta > 0$), and (ii) the more redistribution is expected to characterize the social contract.

Most importantly, it is worth observing that, in contrast with common interpretations of Rawls' Theory, once the optimal scheme of wages and salaries has been achieved, any additional increase in redistribution would not ameliorate the scheme, proving that legitimate inequalities are clearly permitted in the Theory of Justice as Fairness.

4 Optimality under Uncertainty Conditions

4.1 Universal ex-post pareto-efficiency

From the previous Section, it is evident itself that the concordant and the discordant-state imply two different optimal social contracts. More specifically, it can be shown that the optimal scheme of wages and salaries is uniquely defined (i.e. $\beta_1 = \beta_{21} = \beta_{22}$) if and only if $a_H = a_L$ and $\theta_H = \theta_L$, which would render the Difference Principle irrelevant due to the absence of inequalities. Thus, a single optimal social contract is not required to exist, unless valid motivations are adduced by which one or the other state is considered.

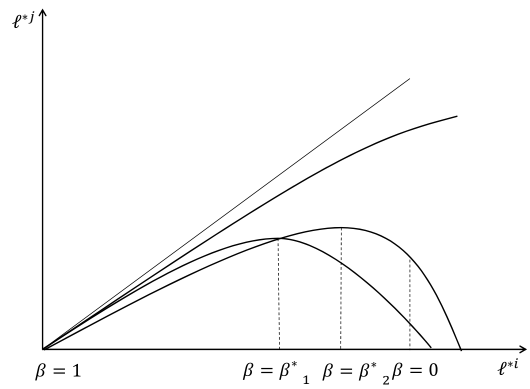
To better emphasize this aspect, let's consider the relationship between the two optimal lifetime income prospects which we will refer to as the 'Rawlsian-efficiency frontier', where the jth individual is assumed to be the least-advantaged in both the concordant and the discordant-state (so, β^*_{21} and β^*_{22} are indicated by β^*_2 hereafter).

To the extent that the lifetime income prospect of the most-advantaged is strictly decreasing with respect β, the optimal lifetime income prospect of the least-advantaged can be plotted with respect to the optimal lifetime income prospect of the most-advantaged, where β must be *decreasing* along the x-axis (Fig. 2[17]). Evidently, if the lifetime income prospect of the least-advantaged individual is $\cap$-shaped with respect to β, then the (state-contingent) Rawlsian-efficiency frontier must be $\cap$-shaped as well. Differently, if the lifetime income prospect of the least-advantaged is strictly decreasing with β, then the corresponding frontier must be positively sloped (Fig. 2). More precisely, to the extent that $\beta^*_1 \geq \beta^*_2 \geq 0$, if $\beta^*_1 = 0$ then $\beta^*_2 = 0$, not vice versa (see the proof of Proposition 3.4); equivalently, if the lifetime income prospect 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 optimal social scheme of wages and salaries cannot be identified independently of the state of the world, a criterion for optimality under uncertainty conditions must be chosen. 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'.

[17] To avoid any misunderstanding, it is worth observing that the Rawlsian-efficiency frontiers depicted in Fig. 2 must not be confused with the OP-curve illustrated by Rawls (2001, p.62) which is, instead, representing the combinations of lifetime income prospects for each possible scheme of wages and salaries.

Figure 2: Rawlsian-efficiency frontiers



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 latter 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 information set at time 0 (i.e., behind the veil of ignorance), the 'ex-ante efficiency' approach would be a non-starting, as both individuals would be clearly associated to the same expected lifetime income prospect as defined with respect to the four equally-probable and mutually-exclusive possible states (i.e., Θ_{HH}, Θ_{HL},Θ_{LH}, Θ_{LL}).[18] Differently, the universal ex-post

[18]More precisely, if both symmetric states of the world are considered with respect to preferences, then each individual has an equal chance to be the one endowed with better

approach is definitely to be preferred for our purposes. By the latter, (state-contingent) lifetime income prospects are not aggregated across different states at the individual level. Instead, an ordering among different schemes is defined by comparing state-contingent income distributions with different degrees of inequality, which is the very scope of the Rawlsian 'Difference Principle'.

Formally, let $\beta^A \in [0,1]$ be the scheme whose corresponding state-contingent distributions of lifetime income prospects are $\bar{\ell}_1^A = \{\ell_{HH}^A, \ell_{LL}^A\} = \{\ell_{LL}^A, \ell_{HH}^A\}$ (concordant-states), and $\bar{\ell}_2^A = \{\ell_{HL}^A, \ell_{LH}^A\} = \{\ell_{LH}^A, \ell_{HL}^A\}$ (discordant-states), with the equivalence conditions holding by symmetry. From the previous Section, given $a_H, a_L, \theta_H, \theta_L$, recall the state-contingent optimal schemes β_1^* and β_2^* for the concordant and discordant-state, respectively, with $\beta_1^* > \beta_2^*$. Hence, it must be the case that β^A is the universally ex-post optimal scheme if there is no $\beta^B \neq \beta^A$ such that, together, (i) $\bar{\ell}_1^B$ is a pareto-improvement of $\bar{\ell}_1^A$, and (ii) $\bar{\ell}_2^B$ is a pareto-improvement of $\bar{\ell}_2^A$.

Proposition 4.1. *The universally ex-post optimal scheme of wages and salaries is:*

$\diamond$ $\beta^* = 0$*, if the lifetime income prospect of the least-advantaged individual is strictly decreasing with respect to* β *in both the concordant and the discordant-state;*

$\diamond$ $0 \leq \beta^* \leq \beta_1^*$*, if the lifetime income prospect of the least-advantaged individual is* $\cap$*-shaped in the concordant-state, whatever the discordant-state.*

Proof. Straightforward from Appendix A.3 and A.4. □

Recall that, by construction of the Rawlsian-efficiency frontiers in Fig. 2, if the lifetime income prospect of the least-advantaged is strictly decreasing in β, then the Rawlsian-efficiency frontier is strictly increasing in $(1-\beta)$; whereas, if the lifetime income prospect of the least-advantaged is $\cap$-shaped in β, then the Rawlsian-efficiency frontier is $\cap$-shaped too $(1-\beta)$. Proposition 4.1 highlights that, if the Rawlsian-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. 2), it must be the case that both individuals are made better off, whatever the state, until $\beta^* = 0$ is attained.

Instead, if the frontier is strictly increasing in the discordant-state only (so, $\cap$-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 lifetime income prospect of the most-advantaged, it must be the case that there exists at least

propensity to effort. Also, to the extent that both individuals have an equal chance to be endowed with a better native talent, four equally probable states are obtained, each one corresponding to a precise distribution of lifetime income prospects. Hence, even if the size of the expected lifetime income prospect is varying with β, the same expected lifetime income prospect is to be associated to both individuals for each β.

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 Rawlsian-efficiency frontiers are both $\cap$-shaped like in Fig. 2, all social contracts such that $\beta_1^* < \beta < 1$ (left-side in Fig. 2) cannot be optimal in that, as in the previous case, the lifetime income prospect 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. 2), optimality is obtained as there are no alternative schemes by which an universal ex-post pareto-improvement can be attained; by reducing β from β_1^*, i.e. increasing the lifetime income prospect 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 Universal ex-post rawls-efficiency

Universal ex-post optimality is supposed to account for the lifetime income of both, the most-advantaged and the least-advantaged individual, in a way that resembles the idea of pareto-dominance. However, even if state-contingency is not even mentioned in Rawls' Theory, one may argue that 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 optimality, as defined in Proposition 4.1, may be weakened according to the maximin principle by focusing exclusively on the least-advantaged individual as follows.

Definition 4.2. *An allocation is said to be universally ex-post rawls-efficient if there is no other feasible allocation by which the lifetime income prospect of the least-advantaged individual can be increased.*

In this view, the definition of the optimal social contract becomes less stringent as compared to the standard universal ex-post efficiency. The following Proposition identifies, according to Definition 4.2, the intervals the optimal scheme must belong to, depending on the shape of the (state-contingent) Rawlsian-efficiency frontier.

Proposition 4.3. *According to the definition of universal ex-post rawls-efficiency, the optimal scheme of wages and salaries is:*

$\diamond$ $\beta^* = 0$*, if the lifetime income prospect of the least-advantaged individual is strictly decreasing with respect to* β *in both the concordant and the discordant-state;*

$\diamond$ $0 \leq \beta^* \leq \beta_1^*$*, if the lifetime income prospect of the least-advantaged individual is strictly decreasing with respect to* β *in the discordant-state but* $\cap$*-shaped in the concordant-state;*

$\diamond$ $\beta_2^* \leq \beta^* \leq \beta_1^*$*, if the lifetime income prospect of the least-advantaged individual is* $\cap$*-shaped in both the concordant and the discordant-state.*

Proof. Straightforward from Appendix A.3 and A.4. □

Although the $\beta^* = 0$ solution is the same as in Proposition 4.1, the $\beta^* > 0$ is now more articulated in that, two different possibilities are to be conceived. More precisely, if the Rawlsian-efficiency frontier is $\cap$-shaped in the concordant state, and strictly increasing in the discordant-state, then it must be the case that all schemes such that $\beta \in]\beta_1^*, 1]$ can be ameliorated according to Definition 4.2 by opting for β_1^*. Moving further to the right from $\beta = \beta_1^*$, to the extent that the frontier is $\cap$-shaped in the concordant-state, there is no alternative scheme by which the lifetime income prospect of the least-advantaged is increased independently of the state; this is similar to the result obtained in Proposition 4.1.

Remarkably, if both frontiers are $\cap$-shaped, then schemes in the interval $\beta \in]\beta_1^*, 1]$ cannot be optimal as before. However, in contrast with Proposition 4.1, the rest of the schemes are not optimal any longer, because schemes in the interval $\beta \in [0, \beta_2^*]$ are universally ex-post rawls-dominated by all schemes in the interval $\beta \in [\beta_2^*, \beta_1^*]$; that is, by moving to the left from $\beta = 0$ to $\beta = \beta_2^*$, the lifetime income prospect of the leat-advantaged increases independently of the state. Evidently, as compared to Proposition 4.1, universally ex-post rawls-optimal social contracts are a subset of the more general universal ex-post case.[19]

From Definition 4.2, *partial justice orderings*[20] can be immediately derived. Formally, let $\ell^j_{LL}(B)$, $\ell^j_{LH}(B)$, $\ell^j_{LL}(A)$ and $\ell^j_{LH}(A)$ be the lifetime income prospect of the (jth) least-advantaged individual as obtained when the schemes 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 4.2,

$$\ell^j_{LL}(B) \gtreqless \ell^j_{LL}(A),\ \ell^j_{LH}(B) \gtreqless \ell^j_{LH}(A) \Longleftrightarrow \beta^B \gtreqless \beta^A;\ \beta^B || \beta^A \ otherwise.$$

Basically, for an 'universal ex-post rawls-improvement' to occur, one of the two schemes of wages and salaries must be enhancing the lifetime income prospect of the least-advantaged in both, the concordant and the discordant-state.

As regards the formal relationship between universal ex-post rawls-improvements and universal ex-post rawls-optimality, it is worth observing that the optimality of a scheme 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

[19] Within a dynamic perspective, the definition of the optimal contract would be sensitive to the initial conditions, as one may end up with differing β's depending on the starting-gate. On the one hand, the benchmark may be defined as the case with no redistribution at all, so that individuals are expected to agree on β_2^*. Conversely, if the benchmark is assumed to be the case of no inequalities (behind the veil of ignorance), then individuals are expected to agree on β_1^*.

[20] 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).

condition for the universal ex-post rawls-improvement to occur.[21]

5 Conclusions

The model we discussed in the previous Section allows to identify the inequalities to be regarded as legitimate according to our interpretation of Rawls' Theory of Justice as Fairness. As such, starting from the original position — where inequalities do not exist by definition — inequalities of lifetime income prospects are permitted to the extent that they stimulate a higher incentive to effort in education and, as a result, economic growth, where the latter must be more than compensating the least-advantaged for being the penalized one. More specifically, inequalities must be tolerated in the society until the major incentive to effort in education (for both the most- and the least-advantaged individual) is more than compensating the least-advantaged for being the penalized one by inequality. In this perspective, the social agreement on the redistributive programme, i.e. the scheme of wages and salaries to be implemented in the labor market, is inevitably affecting the incentives to effort of individuals, so that, in our view, Rawls Theory goes well beyond justice issues in such a way as to accommodate efficiency issues as well.

Even if Rawls' Theory has been largely evoked in the existing literature as the starting-gate of egalitarianism of opportunity, it is worth highlighting that Rawls conceives the redistribution as compensating for different realized talents (earnings ability) where, to the extent that the realized talent is co-determined by individual decisions of effort in education and native abilities, preferences must be inevitably accounted for. Differently, within Roemer's (1993, 1998) ideal of *leveling the playing field*, compensation applies to the final distribution of outcomes in such a way as to compensate the sole disparities originating from different circumstances, i.e., factors beyond individual control. As such, individual decisions are unaffected, and preferences radically disregarded (Roemer and Trannoy 2016). Last but not least, Rawls does not recommend compensation for all circumstances, as differences in terms of native abilities must be preserved in such a way as to fully exploit such differences within the scheme of social cooperation inspiring the well-ordered society.

In contrast with common thinking, we claim that Mirrlees (1971) and Stiglitz (1987), whose models investigate the maximum tolerable redistribution in the presence of efficiency implications, are definitely better references for Rawls'

[21]Clearly, it is not necessary because $\beta^B \succ \beta^A$ may occur even if $\beta^B, \beta^A \notin [\beta_2^*, \beta_1^*]$. In addition, sufficiency does not hold because the optimality of β^B (i.e., $\beta^B \in [\beta_2^*, \beta_1^*]$) and the non-optimality of β^A (i.e., $\beta^A \notin [\beta_2^*, \beta_1^*]$) do not necessarily imply $\beta^B \succ \beta^A$; e.g., let's suppose that (i) $\beta^B \in [\beta_2^*, \beta_1^*]$ and (ii) $\beta^A \in [0, \beta_2^*]$. By (i) and (ii), it must be the case that $\ell^j_{LL}(B) > \ell^j_{LL}(A)$, meaning that, in the concordant-state, the lifetime income prospect of the least-advantaged is higher when β^B is implemented. However, if $\ell^j_{LH}(B) < \ell^j_{LH}(A)$, then β^A is to be preferred in the discordant-state. To the extent that the two schemes of wages and salaries 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.

Theory. However, while Stiglitz and Mirrlees consider the so called 'leisure trade-off' (Musgrave 1974), so that the distortion of labor supply matters, this is not relevant in Rawls' Theory. In the Theory, the expected lifetime income prospect is defined as the product of the earnings ability and the potential time endowment which, evidently, is the same for all individuals. In this sense, leisure is included in the set of primary goods by definition, so that the least-advantaged individual can be identified with respect to the sole earnings ability and, above all, independently of the labor supply. Evidently, in the working life individuals may obtain different income and wealth depending on their labor supply, however, according to Rawls, by virtue of 'background procedural justice', these inequalities are to be regarded as just to the extent that they are obtained in a society ensuring the respect of basic principles.[22]

Finally, the set of potentially optimal social contracts has been determined according to our interpretation of Rawls' Theory. The impossibility of a unique social contract is due to the existence of two possible states with respect to the revelation of native talent which, according to Rawls, occurs ex-post only. As such, two state-contingent income distributions are considered; in the concordant-state the individual endowed with higher propensity to effort is associated ex-post to the better native talent, whereas the other individual is penalized in terms of both propensity to effort and native talent. In the discordant-state, instead, the contingent income distribution is obtained by assuming that the individual with the higher propensity to effort reveals of the worse native talent ex-post, whereas the individual with lower propensity to effort is associated ex-post to better native talent.

In order to identity the set of (potentially) optimal social contracts, the definition of universal ex-post pareto-efficiency has been modified in such a way as to accommodate Rawls' priority for the worse-off; according to the latter, an allocation is said to be optimal if the sole least-advantaged is made better-off whatever the state of the world revealing ex-post. As a result, the set of rawls-optimal social contracts is shown to be a subset of universally ex-post pareto-optimal ones.

[22] "Taking the basic structure as the primary subject enables us to regard distributive justice as a case of pure background procedural justice: when everyone follows the publicly recognized rules of cooperation, the particular distribution that results is acceptable as just whatever that distribution turns out to be" (Rawls 2001, p.54).

Appendix

A.1: Proof of Property 3.1

From (6), by taking the first-order derivatives $\frac{\partial e^{i*}}{\partial e^j}$ and $\frac{\partial e^{j*}}{\partial e^j}$, it can be shown that

$$\begin{aligned} \frac{\partial e^{i*}}{\partial e^j} = 0 \Leftrightarrow e^{i*} = \frac{a_H(-1+\beta)}{4(-1+a_H)} \\ \frac{\partial e^{j*}}{\partial e^j} = 0 \Leftrightarrow e^{j*} = \frac{a_L(-1+\beta)}{4(-1+a_L)} \end{aligned} \tag{9}$$

which are two maximum points.

To prove that the incentive of each individual on the reaction function is determined by the two shares of lifetime income, let $x^i = \frac{E[\ell^i]}{E[\ell^i]+E[\ell^j]}$ and $x^j = \frac{E[\ell^j]}{E[\ell^i]+E[\ell^j]}$ indicate the ith and the jth share of lifetime income as obtained by replacing (1) in (3). By the implicit function theorem

$$\frac{\partial e^j}{\partial e^i} = -\frac{\frac{\partial^2 x^j}{\partial e^j \partial e^i}}{\frac{\partial^2 x^j}{\partial^2 e^j}}$$

where the denominator corresponds to the second-order condition which is negative. As such, given the sign function $S\{\cdot\}$, it must be the case that $S\left\{\frac{\partial e^j}{\partial e^i}\right\} = S\left\{\frac{\partial^2 x^j}{\partial e^j \partial e^i}\right\}$, where

$$\frac{\partial^2 x^j}{\partial e^j \partial e^i} = \frac{(e^j - e^i)(1-\beta)}{(e^i + e^j)^3}$$

This proves that the reaction function has a maximum on the bisectrix, i.e. when $e^i = e^j$. In addition, it must be the case that $S\left\{\frac{\partial^2 x^j}{\partial e^j \partial e^i}\right\} = S\left\{\frac{\partial^2 x^j}{\partial \ell^j \partial \ell^i}\right\}$, where

$$\frac{\partial^2 x^j}{\partial \ell^j \partial \ell^j} = \frac{\ell^j - \ell^i}{(\ell^i + \ell^j)^3}$$

This proves that any increase of ℓ^j (or, equivalently, e^j) is more share increasing when $\ell^j > \ell^i$ (or, equivalently, $e^j > e^i$); i.e., if j is the individual with the highest ℓ, then any increase of ℓ^j is more effective on the jth share when ℓ^j is greater than ℓ^i.

A.2: Proof of Proposition 3.2

The two Nash-equilibria in (7) are straightforward from (6). Let $Max\{e^{*i}\}$ and $Max\{e^{*j}\}$ be, respectively, the ith and jth maximum of the reaction function in (9). Recalling the interior Nash-equilibrium in (7), it must be the case that

$$e^*_H - Max\{e^{*j}\} = \frac{a_L\left(-1 + \frac{4a_H^2(-1+a_L)^2}{(a_H+a_L-2a_Ha_L)^2}\right)(1-\beta)}{4(1-a_L)} > 0$$

$\forall\ a_H, a_L :\ 1 \geq a_H > a_L \geq 0$. This proves that the Nash-equilibrium occurs on the decreasing line of the jth reaction function (see Fig. 1). Also,

$$e_L^* - Max\{e^{*i}\} = \frac{a_H\left(-1 + \frac{4(-1+a_H)^2 a_L^2}{(a_H+a_L-2a_H a_L)^2}\right)(1-\beta)}{4(1-a_H)} < 0$$

$\forall\ a_H, a_L :\ 1 \geq a_H > a_L \geq 0$. This proves that the Nash-equilibrium occurs on the increasing line of the ith reaction function (see Fig. 1).

A.3: Proof of Proposition 3.3

In the concordant-state, let ℓ_{HH} be the (state-contingent) lifetime income prospect of the individual with better native talent and higher propensity to effort, as obtained by replacing (1) and (7) into (3). From (7), to the extent that $e_H^* > e_L^*$ and $\left|\frac{\partial e_H^*}{\partial \beta}\right| > \left|\frac{\partial e_L^*}{\partial \beta}\right|$ $\forall\ \beta$ with $\frac{\partial e_H^*}{\partial \beta}, \frac{\partial e_L^*}{\partial \beta} < 0$, it must be the case that

$$\frac{\partial \ell_{HH}}{\partial \beta} = \frac{1}{2}e_L^*\theta_L + \frac{\beta}{2}\frac{\partial e_L^*}{\partial \beta}\theta_L - \frac{1}{2}e_H^*\theta_H + \left(1 - \frac{\beta}{2}\right)\frac{\partial e_H^*}{\partial \beta}\theta_H < 0$$

which proves that, in the concordant-state, the lifetime income of the most-advantaged is decreasing in β. Differently, by considering the (state-contingent) lifetime income prospect of the individual with better native talent and higher propensity to effort (i.e., ℓ_{LL}), it must be the case that

$$\frac{\partial \ell_{LL}}{\partial \beta} = \frac{1}{2}e_H^*\theta_H + \frac{\beta}{2}\frac{\partial e_H^*}{\partial \beta}\theta_H - \frac{1}{2}e_L^*\theta_L + \left(1 - \frac{\beta}{2}\right)\frac{\partial e_L^*}{\partial \beta}\theta_L \lesseqgtr 0$$

As far as the sign of the latter is not uniquely defined, the (state-contingent) lifetime income of the least-advantaged may be either increasing or decreasing depending on β. More specifically, by considering the first-order condition,

$$\frac{\partial \ell_{LL}}{\partial \beta} = 0 \Leftrightarrow \beta_1^* = \frac{3a_L\theta_L + a_H((-1+a_L)\theta_H - 3a_L\theta_L)}{2(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L))}$$

which can be shown to be a maximum. To complete the proof we need to prove that there exists a set of parameters $\{a_H, a_L, \theta_H, \theta_L\}$ such that $\beta_1^* \in [0, 1]$. Since the denominator of β_1^* is negative by construction, it must be the case that

$$\begin{aligned}&\beta_1^* \geq 0 \Leftrightarrow a_H\theta_H(1-a_L) \geq 3a_L\theta_L(1-a_H)\\ &\beta_1^* \leq 1 \Leftrightarrow \text{by construction}\end{aligned}$$

This proves that, if $a_H\theta_H(1-a_L) \geq 3a_L\theta_L(1-a_H)$, then the lifetime income prospect of the least-advantaged is increasing until β_1^*, while it is decreasing $\forall\ \beta > \beta_1^*$. On the contrary, if $a_H\theta_H(1-a_L) < 3a_L\theta_L(1-a_H)$, $\beta_1^* < 0$, then the lifetime income prospect of the least-advantaged is decreasing for all $\beta \in [0, 1]$.

Similarly, in the discordant-state, the lifetime income prospects of the two individuals can be obtained by substituting (1) and (7) into (3), given that the better endowed in terms of native talent is the individual with the lowest propensity to effort, and vice versa. As such, the two corresponding (state-contingent) lifetime income prospects are

$$\ell_{HL} = \frac{a_H a_L(-1+\beta)(a_H(-2+\beta)\theta_L + a_L(2a_H\theta_L + \beta((-1+a_H)\theta_H - a_H\theta_L)))}{2(a_H + a_L - 2a_H a_L)^2}$$
$$\ell_{LH} = -\frac{a_H a_L(-1+\beta)(a_H\beta\theta_L + a_L((-1+a_H)(-2+\beta)\theta_H - a_H\beta\theta_L))}{2(a_H + a_L - 2a_H a_L)^2}$$

Remarkably, the least-advantaged is not defined *a priori*, so that condition (8) applies. As far as the redistributive and the dis-incentive effect are clearly equally signed for the most-advantaged individual (like in the concordant-state), let's focus on the effect of a marginal increase of β on the least-advantaged. We first suppose that $\Theta_{HL} > \Theta_{LH}$, or, equivalently, $a_H\theta_L(1-a_L) > a_L\theta_H(1-a_H)$, so that ℓ_{LH} is the least-advantaged. Given

$$\frac{\partial \ell_{LH}}{\partial \beta} = \frac{a_H a_L(a_H(1-2\beta)\theta_L + a_L(-(-1+a_H)(-3+2\beta)\theta_H + a_H(-1+2\beta)\theta_L))}{2(a_H + a_L - 2a_H a_L)^2}$$

it can be shown that a maximum is obtained for $\beta = \beta^*_{21}$, where

$$\beta^*_{21} = \frac{a_H\theta_L + a_L(3(-1+a_H)\theta_H - a_H\theta_L)}{2(a_H\theta_L + a_L((-1+a_H)\theta_H - a_H\theta_L))}$$

with

$$\beta^*_{21} \geq 0 \Leftrightarrow a_H\theta_L(1-a_L) \geq 3a_L\theta_H(1-a_H)$$
$$\beta^*_{21} \leq 1 \Leftrightarrow \text{by construction}$$

This proves that, when the individual with better native talent but lower propensity to effort is the least-advantaged (i.e., $\Theta_{HL} > \Theta_{LH}$), if $a_H\theta_L(1-a_L) \geq 3a_L\theta_H(1-a_H)$, then the lifetime income prospect of the least-advantaged is increasing until β^*_{21}, while it is decreasing $\forall\ \beta > \beta^*_{21}$. In contrast, if $a_H\theta_L(1-a_L) < 3a_L\theta_H(1-a_H)$, then the lifetime income prospect of the least-advantaged is decreasing for all $\beta \in [0,1]$.

Finally, let's consider the last possibility, where the least-advantaged individual is the one endowed with worse native talent but higher propensity to effort (i.e., $\Theta_{HL} < \Theta_{LH}$, or, equivalently, $a_H\theta_L(1-a_L) < a_L\theta_H(1-a_H)$). From

$$\frac{\partial \ell_{HL}}{\partial \beta} = \frac{a_H a_L(a_H(-3+2\beta)\theta_L + a_L((-1+a_H)(-1+2\beta)\theta_H + a_H(3-2\beta)\theta_L))}{2(a_H + a_L - 2a_H a_L)^2}$$

it can be shown that a maximum is obtained for $\beta = \beta^*_{22}$, where

$$\beta^*_{22} = \frac{3a_H\theta_L + a_L((-1+a_H)\theta_H - 3a_H\theta_L)}{2(a_H\theta_L + a_L((-1+a_H)\theta_H - a_H\theta_L))}$$

with

$$\beta^*_{22} \geq 0 \Leftrightarrow a_L\theta_H(1-a_H) \geq 3a_H\theta_L(1-a_L)$$
$$\beta^*_{22} \leq 1 \Leftrightarrow \text{by construction}$$

This proves that, when the individual with worse native talent but higher propensity to effort is the least-advantaged (i.e., $\Theta_{HL} < \Theta_{LH}$), if $a_L\theta_H(1-a_H) \geq 3a_H\theta_L(1-a_L)$, then the lifetime income prospect of the least-advantaged is increasing until β^*_{22}, while it is decreasing $\forall\ \beta > \beta^*_{22}$. In contrast, if $a_L\theta_H(1-a_H) < 3a_H\theta_L(1-a_L)$, then the lifetime income prospect of the least-advantaged is decreasing for all $\beta \in [0,1]$.
Finally, considering altogether the conditions for $\beta^*_1, \beta^*_{21}, \beta^*_{22} > 0$, notice that, if the lifetime income prospect of the least-advantaged in the concordant-state is strictly decreasing in β, then it must be the case that the lifetime income prospect of the least-advantaged in the discordant-state is strictly decreasing as well (hint: by contradiction).

A.4: Proof of Proposition 3.4
The definition of the state-contingent optimal schemes of wages and salaries are straightforward from Appendix A.3. To prove that $\beta^*_1 > \beta^*_{21} > \beta^*_{22} > 0$ holds for the (state-contingent) interior solutions, we consider the two differences below, i.e.,

$$\beta^*_1 - \beta^*_{21} = -\frac{(-1+a_H)a_H(-1+a_L)a_L\left(\theta_H^2-\theta_L^2\right)}{(a_H\theta_L + a_L((-1+a_H)\theta_H - a_H\theta_L))(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L))}$$
$$\beta^*_{21} - \beta^*_{22} = \frac{-a_H\theta_L + a_L((-1+a_H)\theta_H + a_H\theta_L)}{a_H\theta_L + a_L((-1+a_H)\theta_H - a_H\theta_L)}$$

As far as both are positive by construction, it must be the case that $\beta^*_1 > \beta^*_{21} > \beta^*_{22}$, which must be also positive by definition of interior solution.

A.5: State-contingent Rawlsian-efficiency frontier
In what follows, we construct the Rawlsian-efficiency frontier for the concordant-state. For the sake of brevity, the same procedure is omitted for the discordant state.

Recall from Appendix A.3 that ℓ_{HH} is strictly decreasing with respect to β. By considering the inverse function, two different solutions are obtained, i.e.,

$$\begin{aligned}\beta = \big(3a_H^2a_L\theta_H - 3a_H^2a_L^2\theta_H - a_Ha_L^2\theta_L + a_H^2a_L^2\theta_L + \\ \pm\sqrt{\big(a_Ha_L\big(-8\big(2a_H(1-2a_L)a_L\ell^*_{HH} + a_L^2\ell^*_{HH} +} \\ + a_H^2\left((1-2a_L)^2\ell^*_{HH} + (-1+a_L)a_L\theta_H\right)\big)\left(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L)\right) + \\ + \ a_Ha_L(a_L\theta_L + a_H(3(-1+a_L)\theta_H - a_L\theta_L))^2\big)\big)\big) \ / \\ (2a_Ha_L(-a_L\theta_L + a_H(\theta_H - a_L\theta_H + a_L\theta_L)))\end{aligned}$$

where, since β must be positively defined, to focus on the case of an interior solution we can consider the sole positive sign in front of the square root.

By replacing β in the lifetime income prospect of the least-advantaged,

$$\begin{aligned}\ell^*_{LL} = \Big(&\big(3a_H^2 a_L\theta_H - 3a_H^2 a_L^2\theta_H + 3a_H a_L^2\theta_L - 3a_H^2 a_L^2\theta_L + \\ &+ \sqrt{\ }\big(a_H a_L\big(-8\big(2a_H(1-2a_L)a_L\ell^*_{HH} + a_L^2\ell^*_{HH} + \\ &+\ a_H^2\big((1-2a_L)^2\ell^*_{HH} + (-1+a_L)a_L\theta_H\big)\big)\big(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L)\big) + \\ &+\ a_H a_L(a_L\theta_L + a_H(3(-1+a_L)\theta_H - a_L\theta_L))^2\big)\big)\big)\big(a_H^2 a_L\theta_H - a_H^2 a_L^2\theta_H + a_H a_L^2\theta_L + \\ &- a_H^2 a_L^2\theta_L + \sqrt{\ }\big(a_H a_L\big(-8\big(2a_H(1-2a_L)a_L\ell^*_{HH} + a_L^2\ell^*_{HH} + a_H^2\big((1-2a_L)^2\ell^*_{HH} + \\ &+\ (-1+a_L)a_L\theta_H)\big)\big(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L)\big) + \\ &+\ a_H a_L(a_L\theta_L + a_H(3(-1+a_L)\theta_H - a_L\theta_L))^2\big)\big)\big)\Big)\Big/ \\ &\big(8a_H a_L(a_H + a_L - 2a_H a_L)^2(a_L\theta_L + a_H((-1+a_L)\theta_H - a_L\theta_L))\big)\end{aligned}$$

which is the Rawlsian-efficiency frontier in the concordant-state.

References

[1] Alexander, S.S. *Social Evaluation Through Notional Choice* The Quarterly Journal of Economics, **88**(4), 597–624, 1974.

[2] Bradford, W. *The economics of Rawlsian justice: can it be neoclassical?* International Journal of Social Economics, **39**(8), 612–623, 2012.

[3] Galor, O., and Zeira, J. *Income Distribution and Macroeconomics.* Review of Economic Studies, **60**(1), 35–52, 1993.

[4] Galor, O., and Moav, O. *From Physical to Human Capital Accumulation: Inequality and the Process of Development.* Review of Economic Studies, **71**(4), 1001–26, 2004.

[5] Hammond, P.J. *Ex-Ante and Ex-Post Welfare Optimality under Uncertainty.* Economica, **48**(191), 235–250, 1981.

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

[7] Kaldor, P.A. *A Model of Economic Growth.* The Economic Journal, **67**(268), 591–624, 1957.

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

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

[10] Rawls, J. *A theory of justice.* Massachusetts: Harvard University Press, 1971.

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

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

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

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

[15] Roemer, J.E. *Equality of opportunity*. Cambridge, MA: Harvard University Press, 1998.

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

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

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