



**HAL**  
open science

# Rawls's original position is not sufficient to specify the rules of cooperations

Bahram Houchmandzadeh

► **To cite this version:**

Bahram Houchmandzadeh. Rawls's original position is not sufficient to specify the rules of cooperations. 2018. hal-01922792v1

**HAL Id: hal-01922792**

**<https://hal.science/hal-01922792v1>**

Preprint submitted on 14 Nov 2018 (v1), last revised 23 Nov 2018 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rawls's "original position" is not sufficient to specify the rules of cooperations.

Bahram Houchmandzadeh  
CNRS, LIPHY, F-38000 Grenoble, France  
Univ. Grenoble Alpes, LIPHY,  
F-38000 Grenoble, France

In his landmark work, "justice as fairness", John Rawls conjectured that the concept of "original position" and "veil of ignorance" will specify the terms of cooperation the individuals forming a society would agree to. We show here that this concept does not carry enough constraints and therefore is not sufficient to reach such an agreement.

## I. INTRODUCTION.

John Rawls, in his landmark work, "Theory of justice" [1, 2], introduced the concepts of "Veil of ignorance" and the "original position" into the contractarianism tradition of political philosophy[3]. He argued that the *veil* strips away all biases from the contractors - due to their sexes, origin, religion,... - during their deliberation (figure 1). Therefore, he conjectured that "the principles of justice the parties would agree to [...] would specify the terms of [their] cooperation"[2, p. 17]. Rawls then developed this idea and concluded that the principle of justice the parties would agree to would be the so called Maximin, a set of rules that maximizes the outcome of the worst-off person of the society.

From its onset, the Maximin deduction came upon criticisms from the *utilitarians* [4, 5] which favor maximizing the total outcome ; various other principles between Maximin and utilitarian can be argued about[6], depending on the weight given to the risk adversity of the rational human.

The purpose of this article is to argue that the debates between these various schools is beside the point : even if the parties, using the veil of ignorance, agree to the same principle of justice (Maximin, utilitarian,...) they cannot agree on a set of rules to achieve the desired principle. In short, the "veil of ignorance" does not remove enough of the original biases to allow for such an agreement.

In order to demonstrate the above statement, I will use a simple mathematical formulation of Rawls theory. Usually, most areas of philosophy are too complex to be modelizable by a mathematical approach. The theory of Rawls however shares many features - such as principles of symmetry, use of initial conditions, optimization procedure, ... - with physical theories. The theory has been restricted (to the most basic social institutions) and abstracted (Rawls insists many times that his view is a theoretical tool and does not reflect any real situation) and all of the simplifying hypothesis have been precisely defined. All these features allows one to put the theory to a mathematical test and this is precisely the purpose of this article.

It is of course not the first time that such an attempt has been made ; the features I have enumerated above have attracted many scientists to mathematical modeling (see [7] for a review) but, to my knowledge, these

works have mostly been dedicated to decide which principles ( Maximin, utilitarian, ...) can be deduced from the "original position" hypothesis.

In the next section, I introduce a simple mathematical modeling of the Rawls theory. This model is restricted to a minimalistic version of Rawls theory, where the outcome for each individual (his quality of life) in the society can be measured by a single number (such as his wealth). I'll show that even in this simplest model, no agreement can be reached and therefore, any more realistic model will only increase the amount of indeterminacy.

The details of mathematical demonstrations are given in the appendices.

## II. MINIMALISTIC RAWLS .

### A. Veil of ignorance.

The minimalist version of Rawls theory I use is the following (figure 1) :  $N$  individuals decide to join and form a society, where the relations between individuals obey a given rule  $\mathcal{R}$  (Slavery, Feudalism, Capitalism, Communism, ...) . In particular, the choice of  $\mathcal{R}$  will decide how wealth and more generally "quality of life" will be distributed in the society. I suppose that this "quality of life" can be measured by a single number. We can assume that when individuals confer on such matters, their selection of the rule  $\mathcal{R}$  is biased by their current position in the present society : A wealthy person will argue for the superiority of capitalism and the sanctity of private property, a feudal lord will be surprised to hear about egalitarian society while a leather worker of 1830 would adhere to the thesis of Fourier, St-Simon, or Proudhon.

Is there a possibility to choose a "*just*" rule  $\mathcal{R}$  ? Can such a thing even exist ? Rawls found an ingenious device to solve this dilemma ("a theory of justice", 1971[1]) in the concept of "Veil of Ignorance". The society the individuals will join is made of vacant positions (for feudalism : few lords, a little more miserable merchants, a large number of totally miserable serfs), and each individual will pick *at random* its new position in the chosen society. In particular, an individual will not know his future age, gender, color, education level. If the chosen society is feudalism and he has picked the position of a female serf, he *will* become an authentic female serf with

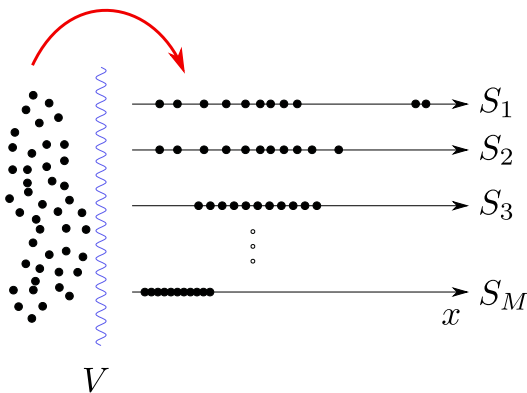


Figure 1. The abstract concept of “social contract” and the formation of society from free individuals. Individuals consent to a rule  $\mathcal{R}_i$  which will result in a type of society  $S_i$  they will join, when a certain kind of resource distribution has been agreed upon. The idea of Rawls is that free individuals are behind a “Veil of ignorance”, not knowing who they are going to be (gender, skin color, education, ...) in the society they will join : they will pick at *random* a place in the society they have chosen.

all the attributes and cognitive capital of such a person in such a society. His brain will be totally rewired and he will lose all memory of his initial position.

Individuals behind the veil will consider all possible societies  $S_1, \dots, S_M$  and choose the *best* one among them.

Does the notion of Veil of ignorance allow one to decide which is best? A short summary of Rawls’s answer is “yes” and the solution is the “Maximin” solution : individuals would not want to be in a very bad position in the new society ; in order to limit such a risk, they will choose an organization  $S_\alpha$  which maximizes the wealth of the least favored person.

The Maximin principle solution is the solution if we consider humans to be totally risk averse. The utilitarians would argue that a rational human should maximize its “expectation” and would therefore choose a society that maximizes the total wealth. Mixing these two strategies (avoiding risk and maximizing expectation) by giving various weight to each factor, depending on our conception of human behavior, will lead to different criteria for choosing the best society.

### B. Indeterminacy of the veil.

The problem with the veil of ignorance is that individuals behind it cannot truly *observe* the possible societies and gauge them. The only thing they can do is to *predict* the kind of society they will create *if* they agree on a set of rules. These predictions are just that : predictions. They depend on the theoretical models in which the individuals believe. Therefore, if the individuals behind the veil have different theoretical models, even if they believe in the same principle of justice, they cannot agree on a

set of rules to achieve this principle.

For example, suppose that all individuals behind the veil adhere to the Maximin as the principle of justice. Some individuals, let us call them neutralist, believe that the total output of the society (or its total wealth) depends mainly on the number of its members and is not too much affected by their inner relations. With this theoretical model in mind, they would conclude that the society with the best Maximin rating would be a totally egalitarian one. Accordingly, they will choose a set of rules  $\mathcal{R}_1$  that is completely redistributive (the mathematical proof of these statements are given in the next section) such as a Marxist one or one with a very progressive taxation scheme.

On the other hand, some individuals believe that inequality among the society members will enhance competition and the total output of wealth. Let us call this belief the “trickle-down” theory. With this theoretical model in mind, they would conclude that the society with the best Maximin rating would be a totally inequalitarian order. Indeed, if inequality creates enough excess wealth (compared to an egalitarian society), the worst-off person in their ideal society would have superior wealth compared to the worst-off person in an egalitarian one. Trickle down people therefore would agree to a set of rules  $\mathcal{R}_2$  that enhances capital accumulation feedback and let 1% of the society (or less) possess most of the total wealth.

Let me stress this point : in the above textbook case, both the neutralist and the “trickle-down” believe in the Maximin principle of justice. However, the rules they would accept to achieve this principle are diametrically opposite.

The veil of ignorance removes all information about the status of an individual in the future, after the veil, but it does not remove his present state of mind and the theoretical models he adheres to. Even if all individuals adhere to the same principle of justice, variability in their inner theoretical models would prevent them to accept the same set of rules.

In the following, using a very simple mathematical modeling, I will demonstrate the above statements. I stress that I consider only very simple models as “proof of principle”. Adding more realism and complexity to the model only enhances the indeterminacy of the Veil.

### III. MATHEMATICAL FORMULATION.

In order to choose a “best” society, we need a comparison tool (a weighting tool) to numerically order the available ones. Let us call  $x$  the abstract measure of quality of life. Von Neumann and Morgenstein [8] introduced such quantification in 1944 when modeling the game theory ; in a more restricted sense,  $x$  would represent the wealth of individuals. A society  $S_i$  is characterized by the value of its vacant positions  $(x_1, x_2, \dots, x_N)$  (figure 1) which we will denote by  $S_i(\mathbf{x})$ . The *function*  $S_i(\mathbf{x})$

captures entirely the distribution of  $x$  for society  $i$ . For example,  $S_1(\mathbf{x}) = \{1, 2, 4\}$  designates a society of 3 individuals where the worst-off one has wealth equal to 1 unit ( $x_1 = 1$ ) while the best-off one has wealth equal to 4 unit, ;  $S_2(\mathbf{x}) = \{0.5, 1, 6.5\}$  designates a less egalitarian society where the wealth of the worst-off person is only 0.5 unit.

Once we have characterized the various societies by their functions, we can choose a criterion  $W$  to ascertain their merit. In other words,  $W[S_i(\mathbf{x})]$  associates a single number (a merit) to society  $i$  with wealth distribution  $S_i(\mathbf{x})$ . For example,  $W_R[S(\mathbf{x})] = \min\{S(\mathbf{x})\}$  uses the wealth of the worst-off person to gauge each society. In the above examples,  $W_R[S_1(\mathbf{x})] = 1$ ,  $W_R[S_2(\mathbf{x})] = 0.5$  and therefore, in the light of *this* merit functional, society  $S_1$  fares better than society  $S_2$ . On the other hand, if we are a utilitarian, we would have chosen a merit function that measures the total wealth of each society  $W_U[S(\mathbf{x})] = \sum x_i$ ; for the above examples,  $W_U[S_1(\mathbf{x})] = 7$  and  $W_U[S_2(\mathbf{x})] = 8$ : by this criterion, the society  $S_2$  is superior to society  $S_1$ .

In mathematics, such a weighting criterion  $W$  is called a *functional* and the art of finding the solution  $S_\alpha(x)$  that maximizes the given functional is called an optimization problem. All areas of fundamental physics are formulated in the optimization framework and the approach is often called a Lagrangian formulation.

Let us come back to our theory of justice. We now have two problems: first we have to choose a weighing functional  $W[]$  and second we have to find the best  $S_\alpha(x)$  corresponding to this functional. What I am going to argue is that even if the veil of ignorance enabled us to choose a particular functional  $W[]$ , there is not enough information to find a “best” solution.

The problem I am considering is a simple numerical one where all the complexity of human behavior have been simplified into some collection of number. What I am going to show is that even in this very simplified framework, there is no unique solution. Therefore, we should not hope for a unique solution in the much more complex domain of human behavior.

We will consider below few simple cases and show that people adhering to the same principle of justice but with different theoretical model of wealth production will choose very different organization of the society.

### A. Rawlsian individuals.

Let us consider a society formed of only two individuals, which makes the demonstration particularly easy. Demonstrations for the general cases can be found in the appendix. I suppose first that all individuals behind the veil are Rawlsian and adhere to the Maximin principle. There are two position  $x_1$  and  $x_2$  available in the society, where  $x_1$  designates the wealth of the worst-off and  $x_2$  the wealth of the best-off person, hence,  $x_1 \leq x_2$ .

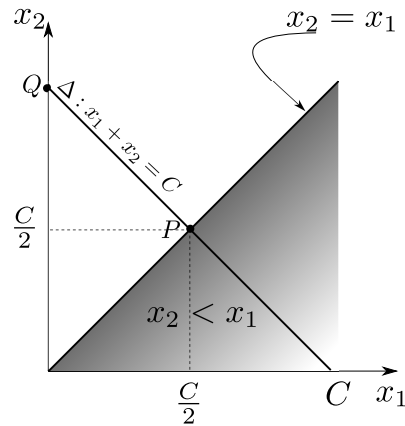


Figure 2. Graphical solution of equations (1-3). For any valid solution  $\{x_1, x_2\}$ , we must have  $x_1 + x_2 = C$  and therefore the solution must reside on the anti-diagonal line  $\Delta$ ; On the other hand, the constraint  $x_1 \leq x_2$  constrains the solution to be above the diagonal  $x_1 = x_2$  (the forbidden area  $x_2 < x_1$  is shown in gray). The best choice for the solution is therefore the point  $P$  where  $x_1 = x_2 = C/2$ . The worst choice is the point  $Q$  corresponding to a totally non egalitarian society.

*Case 1. Neutralist theory.* The theoretical model of the individuals in this case is that the total output  $T$  of the society, whatever the later distribution between its members, is a constant, proportional to the size of society and productivity per person:  $T = x_1 + x_2 = C$ . To find a society  $S(\mathbf{x}) = \{x_1, x_2\}$  that maximizes  $x_1$  is, in mathematical term, finding  $\{x_1, x_2\}$  such that

$$W_R[\{x_1, x_2\}] = x_1 \text{ is maximum} \quad (1)$$

$$x_1 + x_2 = C \quad (2)$$

$$x_1 \leq x_2 \quad (3)$$

This is an optimization problem (equation 1) with two constraints (equations 2-3). Figure 2 demonstrates graphically that the best society is one in which

$$x_1 = x_2 = \frac{C}{2}$$

*i.e.* a totally egalitarian society. The demonstration generalizes to  $N$  individuals (appendix A).

*Case 2. trickle-down theory.* The individuals in this category believe that some output  $T$  of the society is enhanced by some amount of inequality. Let us consider the very simple model where the total output is the same constant  $C$  as before, plus a simple measure of inequality:

$$T = C + \alpha(x_2 - x_1)$$

where  $\alpha$  is a coefficient weighting the importance given to inequality as the enhancer of wealth production. To find the best society, we must find this time  $\{x_1, x_2\}$  such that

$$W_R[\{x_1, x_2\}] = x_1 \text{ is maximum} \quad (4)$$

$$x_1 + x_2 = C + \alpha(x_2 - x_1) \quad (5)$$

$$x_1 \leq x_2 \quad (6)$$

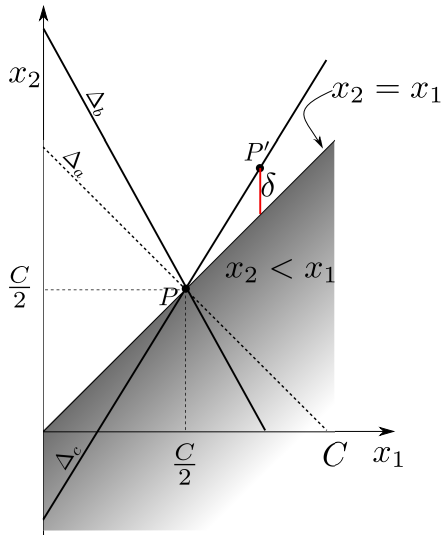


Figure 3. Graphical solution of equations (4-6). The method of resolution is similar to figure (2). Any valid solution  $\{x_1, x_2\}$  must lie on the line  $\Delta_\alpha$  where  $(1+\alpha)x_1 + (1-\alpha)x_2 = C$  (equation 5) ; The constraint  $x_1 \leq x_2$  constrains the solution to be above the diagonal  $x_1 = x_2$ . Four lines  $\Delta_a$  ( $\alpha = 0$ ),  $\Delta_b$  ( $\alpha \in [0, 1]$ ),  $\Delta_c$  ( $\alpha > 1$ ) and  $\Delta_d$  ( $\alpha < 0$ ) are shown on the figure, designating the class of models individuals can believe in.  $\alpha = 0$  corresponds to the neutralist theory investigated before.  $\alpha > 0$  corresponds to the trickle down theory. If  $\alpha \in [0, 1]$ , the slope of  $\Delta_\alpha$  is negative and the best choice is the point  $P$  (similar to the previous case) where  $x_1 = x_2 = C/2$ . If on the other hand, the belief is in  $\alpha > 1$ , the slope of  $\Delta_\alpha$  becomes positive and the point  $P$  becomes the *worst* solution. Any point  $P'$  on the line  $\Delta_\alpha$  with  $x_2 > x_1$  is a better choice, and higher the inequality (the segment  $\delta$ ), better the solution.

Figure 3 shows graphically the result of such optimization with constraint. We see that for a theoretical model where  $0 < \alpha < 1$ , a Rawlsian person will still adhere to an egalitarian society. However, if the Rawlsian believes in a theoretical model of wealth production where  $\alpha > 1$ , then he would advocate a society where inequality is as high as possible.

### B. Utilitarian.

We saw in the above section that a Rawlsian individual will adhere to very different rules of wealth distribution based on his theoretical model of wealth production. There is nothing particular to the Maximin principle and a utilitarian will have the same indeterminacy. Let us revisit the above example from the point of view of a utilitarian person that tries to maximize the total wealth. If he is a neutralist, the graphical solution of figure 4 shows that any wealth distribution  $\{x_1, x_2\}$  is good for him, as long as  $x_1 + x_2 = C$ . On the other hand, if he believes that inequality *reduces* wealth production, *i.e.* the case  $\alpha < 0$  in figure 4, then he would conclude that the best

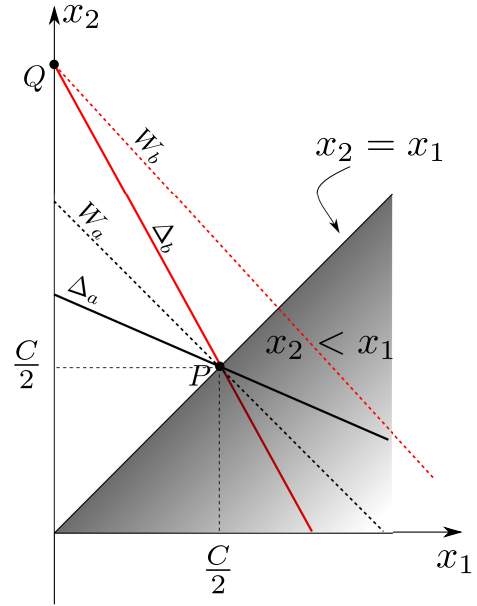


Figure 4. Graphical solution of the optimization problem for a *utilitarian*. The method is similar to the previous cases. The solution  $\{x_1, x_2\}$  must belong to the line  $\Delta_\alpha$  where  $(1+\alpha)x_1 + (1-\alpha)x_2 = C$  where  $x_2 > x_1$  (gray area excluded). The utilitarian maximizes the total wealth  $W = x_1 + x_2$ . If  $\alpha > 0$  (line  $\Delta_b$ , belief in trickle-down theory), the optimum is reached by the point  $Q$  ( $x_1 = 0, x_2 = C/(1-\alpha)$  total inequality) ; if  $\alpha < 0$  (belief in counter productive effect of inequality),  $W$  is maximized for the point  $P$  ( $x_1 = x_2 = C/2$ ). For  $\alpha = 0$ , the slope of  $\Delta_\alpha$  is -1 and any point  $x_1 + x_2 = C$  is a solution. For trickle-down belief, I have considered only the case  $\alpha < 1$  in this figure. The case  $\alpha > 1$  is similar to the one in figure 3 and is not shown here.

society is one where there is total equality. If he believes in trickle-down theory with  $0 < \alpha < 1$ , he will choose a totally unequal society.

## IV. CONCLUSION AND DISCUSSION

In mathematics, when  $n$  unknown quantities are constrained by  $m$  relations and  $m < n$ , the problem is in general *indeterminate* and has no unique solution[9]. I argued in this article that the “veil of ignorance”, as valuable as it is, does not bring enough constraints to determine uniquely the set of rules individuals can agree to. Indeed, even when people agree on a principle of justice (such as Maximin), they can choose diametrically opposing set of rules for the organization of the society. Because Rawls had stated his theory in very precise and falsifiable terms, it was possible to use basic tools of mathematics to demonstrate the above statement.

The above demonstration can be summarized as the following : Individuals behind the veil of ignorance hold different opinions and theoretical models that are not smoothed enough by the veil and lead to failure in reaching an agreement if the original diversity of opinion is

large. This is true even if they accept the principle of justice. Mathematically, we showed that given a principle of justice  $W[]$  which can weight possible societies  $S_\alpha$  ( $W[S_\alpha]$  is the numerical *merit* of the society  $S_\alpha$ ), no agreement can be reached on the set of rules  $\mathcal{R}$  that would organize the society.

In this article, I restricted the discussion to the (un)determination of  $\mathcal{R}$  given the principle of justice  $W[]$ . But can an agreement be reached on the principle of justice  $W[]$ ? The problem is very similar to the problem treated here (it is an optimization problem) but at another abstraction level. An agreement on  $W[]$  can be reached if the individuals have the same theoretical model of human behavior. Humans have a bias in evaluating positive and negative outcome, which has been intensely investigated by social psychologists[10, 11]. If we suppose that individuals are totally risk averse and give much more weight to negative outcome than to positive one, we would naturally choose the Maximin principle, as Rawls did. If we suppose that individuals are neutral in this respect, then we would choose a principle that maximize the mathematical expectation, *i.e.* the total output, as utilitarians do. But variations in individuals risk assessment, which are widespread and may have partially genetic roots[12] shall prevent these individuals to reach an agreement even in the principle of justice.

Mathematically, we could address both these problems by a meta-optimization method : given the variability in the theoretical models of individuals present behind the veil of ignorance, we can compute a solution that minimizes the total dissatisfaction of individuals resulting from the gap between their desired solution and the chosen solution. But then, the individuals have to reach an agreement about how to weight the dissatisfaction ! It seems that there is no rational solution to the multi-level agreement conundrum.

## Appendix A: The optimization problem for arbitrary number of individuals.

### 1. Neutral (Constant output) hypothesis.

Consider a society of  $N$  individuals with vacant positions  $\{x_1, x_2, \dots, x_N\}$  where

$$x_1 < x_2 < \dots < x_N \quad (\text{A1})$$

In the case of constant output hypothesis,

$$\sum_{i=1}^N x_i = C \quad (\text{A2})$$

where  $C = N\mu$  and  $\mu$  is the productivity per person. Let us change our variables from  $x_i$  to  $\delta_i$  :

$$\delta_1^2 = x_1 \quad (\text{A3})$$

$$\delta_i^2 = x_i - x_{i-1} \quad 1 < i \leq N \quad (\text{A4})$$

The set  $\{x_1, \dots, x_N\}$  and  $\{\delta_1, \dots, \delta_N\}$  are equivalent and knowing one leads to knowing the other, *i.e.*

$$x_1 = \delta_1^2$$

$$x_n = \sum_{i=1}^n \delta_i^2$$

This change of variables automatically satisfies constraint (A1), and the constraints (A2) becomes, in the new variable,

$$\sum_{i=1}^N (N - i + 1) \delta_i^2 = C \quad (\text{A5})$$

In order to maximize

$$W(x_1, x_2, \dots, x_n) = x_1 = \delta_1^2$$

subject to constraints (A1A2), I use the Lagrange multipliers formalism and optimize

$$W_\ell = \delta_1^2 - \lambda \left( \sum_{i=1}^N (N - i + 1) \delta_i^2 - C \right)$$

We must therefore solve

$$\frac{\partial W_\ell}{\partial \delta_1} = 2(1 - N\lambda)\delta_1 = 0 \quad (\text{A6})$$

$$\frac{\partial W_\ell}{\partial \delta_i} = 2(N - i + 1)\lambda\delta_i = 0 \quad (\text{A7})$$

The system (A6,A7,A5) has two solutions :

$$\lambda = \frac{1}{N}; \delta_1^2 = \frac{C}{N}; \delta_{i>1} = 0 \quad (\text{A8})$$

$$\lambda = 0; \delta_1 = 0; \sum_{i=2}^N (N - i + 1) \delta_i^2 = C \quad (\text{A9})$$

As  $C/N > 0$ , solution (A8) corresponds to a maximum (point  $P$  on figure 2) while solution (A9) corresponds to a minimum (point  $Q$  on figure 2).

In an ‘‘constant output’’ model, the Rawlian principle leads to an egalitarian society.

### 2. Trickle-down hypothesis.

Before investigating this case, note that equation (A5) corresponds to an ellipsoid in  $N$  dimensions and we could have reached the same conclusion as above by purely geometrical considerations. For simplicity and without loss of generality, we consider only the region where  $\delta_i \geq 0$ . Let us suppose that the point  $P = (\delta_1^*, \delta_2^*, \dots, \delta_N^*)$  belong to the ellipsoid and consider the point  $P' = P + dP = P + (d\delta_1, \dots, d\delta_N)$ . If  $P'$  also belongs to the ellipsoid, we must have

$$\sum_{i=1}^N \kappa_i \delta_i^* d\delta_i = 0$$

where  $\kappa_i = (N - i + 1)$ . To have  $d\delta_1 > 0$  (increase the *merit* of the point  $P$ ), as  $\kappa_i > 0$  and  $\delta_i^* \geq 0$ , we must have at least for one  $j$  :  $\delta_j^* > 0$  and  $d\delta_j < 0$  : the only way to increase the wealth of the worst-off position, we must reduce the inequality. When all  $\delta_{i>1}^* = 0$ , there is no more possibility for improvement.

Consider now the trickle-down hypothesis where the total wealth is a function of the inequality :

$$T = \sum_{i=1}^N x_i = C + \alpha(x_N - x_1)$$

proceeding to the same change of variable (A3-A4) as above, we know have

$$\sum_{i=1}^N (N - i + 1) \delta_i^2 = C + \alpha \sum_{i=2}^N \delta_i^2 \quad (\text{A10})$$

Relation (A10) is again the equation of a conic

$$N\delta_1^2 + \sum_{i=2}^N \kappa_i \delta_i^2 = C$$

where

$$\kappa_i = N - i + 1 - \alpha$$

However, this time the conic is not necessarily an ellipsoid. For example, for  $1 < \alpha < 2$ ,  $\kappa_i > 0$  for  $i = 2, \dots, N - 1$  but  $\kappa_N = (1 - \alpha) < 0$  ! Repeating the above argument, we see that if the point  $P = (\delta_1^*, \delta_2^*, \dots, \delta_N^*)$  belongs to the conic, we can improve the merit by having  $d\delta_1 > 0$  and  $d\delta_N > 0$  : increasing inequality ( $d\delta_N > 0$ ) improves the value of the worst-off position ( $d\delta_1 > 0$ ).

In the trickle-down hypothesis, a Rawlsian person will choose a very unequal society.

## Appendix B: Algebraic vs differential equations.

We considered above only very simple economic models, where algebraic equations were used to capture the wealth of the society. Models can be made complicated at will without changing the result about indeterminacy of the best rule.

For example, the above approach can be criticized as it equates wealth (a stock) and income (a flux). A better approach would be to use differential equations to capture wealth *accumulation*. Let us suppose again that we have two individuals and the surplus output (per year for example) is of the form

$$T = C + \alpha(x_2 - x_1)$$

and this output is shared by the society members ( and added to their wealth ) in proportion to their current

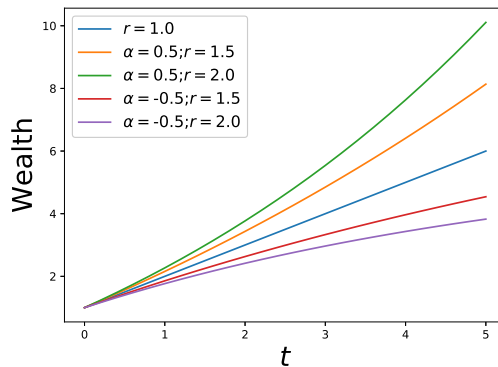


Figure 5. Solution of equation (B4) for various value of initial inequality  $r$  and competition parameter  $\alpha$

wealth  $x_i$ . Then the rate of change of the wealth of each member is

$$\frac{dx_1}{dt} = \frac{x_1}{x_1 + x_2} (C + \alpha(x_2 - x_1)) \quad (\text{B1})$$

$$\frac{dx_2}{dt} = \frac{x_2}{x_1 + x_2} (C + \alpha(x_2 - x_1)) \quad (\text{B2})$$

Let us call  $r$  the ratio of the initial wealth of the two individuals

$$r = x_2(0)/x_1(0)$$

then the above differential equations are easily transformed into

$$x_2(t) = rx_1(t) \quad (\text{B3})$$

$$\frac{dx_1}{dt} = \frac{\alpha(r-1)}{r+1} x_1 + \frac{C}{r+1} \quad (\text{B4})$$

relation (3) states that the two individuals will keep their initial wealth ratio during time. The second equation is solved by

$$x_1(t) = \left( x_1(0) + \frac{C}{\alpha(r-1)} \right) e^{\frac{\alpha(r-1)}{r+1} t} - \frac{C}{\alpha(r-1)} \quad \text{if } r \neq 1$$

$$x_1(t) = x_1(0) + \frac{C}{2} t \quad \text{if } r = 1$$

A utilitarian tries to maximize the total wealth  $x_1(t) + x_2(t) = (1+r)x_1(t)$ . From this solution, we can see for example that if he believes in the deleterious effect of inequality on productivity ( $\alpha < 0$ ) he would adhere to a perfectly egalitarian society ( $r = 1$ ) ; if he hold the contrary belief, he would prefer a society with  $r$  as high as possible (figure 5) . If he is a neutralist ( $\alpha = 0$ ), all inequality values  $r$  are equivalent. Other rules are selected for example if we change the output sharing mechanism between members.

- 
- [1] John Rawls. *A Theory of Justice*. Harvard University Press, Cambridge, Mass, new ed edition, March 2005.
- [2] John Rawls. *Justice as Fairness – A Restatement*. Harvard University Press, Cambridge, Mass, 2rev ed edition, 2001.
- [3] Russ Shafer-Landau. *Ethical Theory: An Anthology*. John Wiley & Sons, Chichester, West Sussex ; Malden, MA, 2nd edition edition, 2012.
- [4] John C. Harsanyi. Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. *American Political Science Review*, 69(02):594–606, June 1975.
- [5] Olatunji A. Oyeshile. A Critique of the Maximin Principle in Rawls ' Theory of Justice. *Humanity & Social Sciences Journal*, 3(1):65–69, 2008.
- [6] Michal Wiktor Krawczyk. A model of procedural and distributive fairness. *Theory and Decision*, 70(1):111–128, January 2011.
- [7] Anthony Laden. Games, Fairness, and Rawls's A Theory of Justice. *Philosophy & Public Affairs*, 20(3):189–222, 1991.
- [8] John Von Neumann, Oskar Morgenstern, Harold William Kuhn, and Ariel Rubinstein. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, N.J. ; Woodstock, 60th anniversary commemorative edition, 2007.
- [9] In most cases, it has an infinite number of solutions.
- [10] Guido Peeters and Janusz Czapinski. Positive-Negative Asymmetry in Evaluations: The Distinction Between Affective and Informational Negativity Effects. *European Review of Social Psychology*, 1(1):33–60, January 1990.
- [11] Michael Siegrist and George Cvetkovich. Better Negative than Positive? Evidence of a Bias for Negative Information about Possible Health Dangers. *Risk Analysis*, 21(1):199–206, February 2001.
- [12] David Cesarini, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. Genetic Variation in Preferences for Giving and Risk Taking. *The Quarterly Journal of Economics*, 124(2):809–842, May 2009.