26 December 2008

# THE RATIONAL, RECURSIVE ORIGINAL POSITION

***Serge-Christophe KOLM****

*Abstract*

Theories of the original position are among the main present-day social ethical theories. Rawls's invalidates utilitarianism whereas Harsanyi's proves it. Harsanyi's focuses on building an impartial evaluation. However, the evaluation of an individual in the original position depends on her preferences about being the various individuals and on her risk-aversion. Yet these different individual evaluations are more alike than individuals' utilities. Consistency demands facing this multiplicity with a further original position, and so on in an infinite regress converging to full unanimity. The outcome is a particular welfarist but non-utilitarian social ethical function.

*Keywords*: original position, Harsanyi, being preferences.

*JEL classification number*: D60

## 1. The original position theory, its problem and the solution

Theories of the original position constitute a main body of the social ethical theories of our times. John Rawls (1971) introduced such a theory in order to displace utilitarianism (and, more generally, welfarism)[1] which was prevalent in his environment. John Harsanyi (1953, 1955, especially 1976) produced another theory of the original position that he and many other scholars consider to be a proof of utilitarianism, on the contrary, and often the basic (or only) proof. A theory of the original position is a social ethical theory that considers that what should be done in society consists in the opinion of the individuals when they do not know which specific individuals they will be (i.e., "in the original position", "behind the veil be ignorance"). The basic intention is that this choice be impartial – not favorable to the self-

* EHESS, IDEP, CREM.

[1] Welfarism is taken to mean the maximization of a classical social welfare function, function of individual's utilities. The term is due to John Hicks (1959). See Rawls (1982) and section 6.2.

interest of some people to the detriment of that of others–, while remaining the choice of the individuals.

However, these theories have different – indeed opposite – conclusions because they differ by having different objectives. Harsanyi's is a theory of impartiality and this only. In contrast, Rawls embeds this objective in a much larger aim. He says that his consideration of an original position is but a moment in his "reflective equilibrium" testing his "considered intuitions" about the principles of justice. This test consists in that there exists an imaginable uncertainty for the individuals in the original position that lead them to choose the principles. This uncertainty can be very large, far beyond the ignorance of which of the actual individuals in the actual circumstances one will be, about circumstances also (and presumably including the possibility of being still other individuals). Rawls's principles of justice undoubtedly pass this test. With sufficiently serious risks, indeed, the individuals in the original position will certainly choose these maximally protective principles: the basic rights and liberty, non-discrimination, and a maximin in "primary goods" (one of which is income or wealth). These individuals "in the original position" are behind this "thick veil of ignorance". This test does not constitute a deductive theory.

In contrast, Harsanyi deals with the problem of impartiality only. The uncertainty in the original position is that which is necessary and sufficient for this purpose: the individuals do not know which of the actual individuals they will be in the actual situation, and a priori this only. This is the "thin veil of ignorance" and the theory is deductive. This quest for impartiality moreover rightfully leads Harsanyi to consider that each individual in the original position has an equal chance to become any of the actual individuals (full uncertainty plus the Condorcet-Laplace axiom of probabilities of the "principle of insufficient reason" would give the same result). This is the setting considered here.

However, in order to evaluate the actual social situation, an individual in the original position needs to have preferences about two things. First, she should have preferences about being one individual or the other, her "being preferences" (or "ontological preferences" in Greek). Second, she should have preferences about uncertainty, a risk-aversion. The being preferences compare being various individuals in all respects, taking into account comparisons of consumption, tastes and preferences, social situations, possibilities and liberty, rights, personal beauty and intelligence, fame, character, information, and so on.

People commonly express various being preferences. All preferences need only be represented by orderings and are assumed to be representable by ordinal utility functions. However, the classical rational evaluations of risk use von Neuman-Morgenstern (VNM) cardinal specifications of ordinal utilities. This behavior in risk is necessary for Harsanyi in order to obtain, with a mathematical expectation, the additive form of a social welfare function that he assimilates to utilitarianism, but it is not for the following considerations. It will nevertheless be assumed for reasons of simplicity in presentation and comparison with Harsanyi (moreover, it may be argued that an ethical theory has better be based on rational behavior).

An individual in the original position has preferences about social states which are her preferences about them given that she faces the a priori uncertainty that she could be any of the actual individuals with probability $1/n$, where $n$ is the number of individuals in the society. This depends on her being preferences and risk-aversion. Therefore, each pair of being preferences and risk-aversion gives, in the original position, a preference ordering of the social states. Which one should we choose? Since actual individuals have a priori different being preferences and risk-aversions, using their pairs of these preferences provides $n$ a priori different evaluations in the original position, one for each actual individual. Now, Harsanyi assumes that the individuals in the original position have the same preferences over the social states. This implies that they have the same being preferences and risk-aversion, which is counterfactual. If we chose these preferences of some impartial external observer, who is she and what are they? The only preferences that exist in society are those of the actual individuals.

However, even though the individuals' preferences in the original position do not present, a priori, the identity or unanimity assumed by Harsanyi (or prefer the same state as assumed by Rawls), a progress has nevertheless been made in this direction. Indeed, each individual evaluation of a social state from the original position is an increasing function of all the individuals' utilities for this social state. Hence, if two alternative social states are ordered similarly by every actual individual's preferences, one gives a higher value than the other (or the same value) to every individual utility function, and the same happens to every individual's evaluation function in the original position. Therefore, the set of pairs of social states that are unanimously ordered by all individuals looses no element when one passes from individual's actual preferences to their preferences in the original position. In general, this set changes, and, therefore, it expands. This means that the individuals agree more. A consequence is that the corresponding Pareto set gains no element and, in general, shrinks.[2]

Nevertheless, we are still left, in the original position, with *n* a priori different preference orderings of the social states and corresponding evaluation functions, one for each individual. Consistency seems to require one to deal with this problem as one just dealt with the formally identical initial one created by the multiplicity of the different actual individual preferences about social states. Especially since this method has led to some progress towards a uniformization of preferences. That is to say, one has to consider an original position of the original position. However, the same problem is faced: the choice of being preferences and risk-aversion for application to being the various individuals in the original position. The same discussion and solution gives another set of *n* evaluation functions in this second-order original position, each using the being and risk preferences of each individual for deriving this second-order function from the ones in the first-order (standard) original position. For the same reason and in the same sense as with the comparison between individuals' utilities and their first-order original-position evaluations, the set of the *n* second-order original-position evaluations agree more between them when ordering the actual social states than the first-order ones do: the set of pairs of social states unanimously ordered looses no element and generally expands, and the Pareto set gains no element and generally shrinks.

Then, the problem of having *n* a priori different evaluation functions in the second-order original position has to be dealt with in the same way again, by considering a third-order original position, and so on. At each step, the orderings of social states implied by the individuals' evaluation functions become more alike: the set of unanimously compared pairs of actual social states looses no element and generally expands, and the Pareto set gains no element and generally shrinks. In normal conditions, when the order of the original position tends to infinity, the corresponding individual orderings of the actual social states converge towards the same ordering. The corresponding evaluation function is the "social welfare function" produced by this theory of the "recursive" or "infinite-regress original position". The explanations provided show that this theory can claim to be the complete, or rational, theory of the original position.

---

[2] The appendix provides full, precise and explicit statements and proofs.

By construction, this social evaluation function is an increasing function of individuals' utility functions. It is not utilitarianism, however, and not even a sum of increasing functions depending each of an individual's utility meaning happiness that Harsanyi and others see as utilitarianism (the first-order original-position VNM evaluation functions only have this structure, but there are a priori $n$ of them).

**2. The general model**

There are $n$ individuals indexed by $i$. They have preferences about a social state in the classical sense denoted as $x$. This social state encompasses all that concerns them (in particular their allocations of all commodities, incomes, rights and liberties, public goods, etc.) – except their relevant preferences as usual. The social choice is that of such a $x$ directly or indirectly (such as by various policies, rules, laws or principles). We consider functions of $x$ that represent preference orderings of the possible $x$.

Individual $i$ has an ordinal utility function of $x$, and $u_i(x)$ is a specification of this ordinal function. However, individual $i$ also has being (ontological) preferences. Denote as $(b_j, x)$ "being individual $j$ when state $x$ holds". Individual $i$ has, a priori, preferences over such pairs. These preferences are represented by an ordinal utility function and $\tilde{u}_i(b_j, x) \in \rho_i \subset \Re^1$ denotes a specification of this function, where $\rho_i$ is a compact subset of $\Re^1$. The preferences on $x$ represented by the function $\tilde{u}_i(b_i, x)$ are tautologically represented by the function $u_i(x)$. Since the specification $u_i$ of an ordinal utility functions is a priori arbitrary, one can choose, as specification $u_i$, $u_i(x) = \tilde{u}_i(b_i, x)$, for each $x$ and $i$.

Being individual $j$ implies having all that concerns this person including allocation, social situation and relations, and so on – described in $x$ –, and her personal mental and physical characteristics, including her preferences over $x$. Therefore, the orderings of $x$ by $\tilde{u}_i(b_j, x)$ and by $u_j(x)$ are the same. These two functions are specifications of the same ordinal function of $x$. Therefore, there exists a $\Re^1 \to \Re^1$ increasing function $h_{ij}$ such that (with the usual notation for the composition of functions),

$\tilde{u}_i(b_j, x) = h_{ij}[u_j(x)] = h_{ij} \circ u_j(x)$,[3]

which is a general form for all $i$ and $j$ by writing $h_{ii}=1$.

Let $g_i(\tilde{u}_i)$ denote a specification of individual $i$'s VNM cardinal utility. Function $g_i$ is increasing (and, from cardinality, it can be replaced by any $a_i g_i + b_i$ where $a_i$ and $b_i$ are constant and $a_i>0$). Denote $v_i(x) = g_i \circ u_i(x)$.

## 3. The multiple original position

In the original position, individual $i$ faces the risky prospect of becoming each individual with an equal probability $1/n$ (as a condition of justice and fairness). Her evaluation of this prospect is her expected utility

$$v_i^1(x) = n^{-1}\Sigma_j g_i \circ \tilde{u}_i(b_j, x) = n^{-1}\Sigma_j g_i \circ h_{ij} \circ u_j(x) = n^{-1}\Sigma_j \gamma_{ij} \circ u_j(x)$$

$$= n^{-1}\Sigma_j g_i \circ h_{ij} \circ g_j^{-1} \circ v_j(x) = n^{-1}\Sigma_j H_{ij} \circ v_j(x), \quad (1)$$

where $\gamma_{ij} = g_i \circ h_{ij}$ and $H_{ij} = g_i \circ h_{ij} \circ g_j^{-1} = \gamma_{ij} \circ g_j^{-1}$ are increasing functions.

In particular, $H_{ii}=1$.

These functions represent the $n$ orderings of the $x$ of the individuals in the original position. They are a priori different. However, the orderings of the $x$ defined by the $v_i^1$ are generally more alike than those defined by the $v_i$.

Indeed, since the functions $\gamma_{ij}$ or $H_{ij}$ are increasing for all $i$ and $j$, equations (1) show that the $v_i^1$ for all $i$ are increasing functions of the $u_j$ or $v_j$ for all $j$, for each $x$. Then, when comparing two social states $x$, passing from the orderings defined by the utility functions $u_i$ or $v_i$ to those in the original position defined by functions $v_i^1$ maintains the unanimous

---

[3] In order to be on the safe side, let us point out that one cannot in general take the same specification of these ordinal functions, $\tilde{u}_i(b_j, x) = u_j(x)$ for all $i$, $j$, because the $\tilde{u}_i$ are also functions of the $b_j$ (being preferences). For instance, if $\tilde{u}_i(b_j, x) = \tilde{u}_k(b_j, x) = u_j(x)$ and $\tilde{u}_i(b_\ell, x) = \tilde{u}_k(b_\ell, x) = u_\ell(x)$, this contradicts the fact that one can have $\tilde{u}_i(b_j, x) > \tilde{u}_i(b_\ell, x)$ and $\tilde{u}_k(b_\ell, x) > \tilde{u}_k(b_j, x)$, that is, for social state $x$ individual $i$ prefers to be individual $j$ than individual $\ell$, and individual $k$ has the reverse preference.

comparisons by indifference, strict preference, preference or indifference, and the latter plus strict preference for at least one *i* (Pareto domination, which becomes, actually, unanimous strict preference) – see the appendix. Hence, the sets of pairs related by each type of these unanimous preferences or by the Pareto domination loose no element. However, they generally change. Therefore, they generally expand. This passage to the original position results, in general, in adding new unanimous or Pareto pairwise comparisons while loosing none. As a result, in particular, since a state Pareto-efficient in the original position (with the $v_i^1$) is not Pareto-dominated by any other possible state with these preferences (by definition), it is not Pareto-dominated by a possible state with the $v_i$ and $u_i$ either, and therefore it is Pareto-efficient (with the latter, actual preferences). Hence, when passing to the original position the Pareto set gains no element, and it generally shrinks.

**4. The moral regress of original positions**

Moreover, the problem of having one evaluation per individual in the original position is analogous to the initial problem of having various individual evaluations in the real world. Consistency suggests or requires facing this problem with the same method, especially since it led to some progress. We therefore have to consider an original position of the original position where the individuals face the risk of having each of the evaluations $v_i^1$ with the same probability $1/n$. Then, individual *i*'s evaluation in this second-degree original position, $v_i^2(x)$, obtains from the $v_j^1(x)$ as the latter obtained from the $v_k(x)$, that is,

$$v_i^2(x) = n^{-1}\Sigma_j H_{ij} \circ v_j^1(x) .$$

There still are *n* evaluations. However, for the same reason and in the same sense as above, the orderings of the *x* defined by the $v_i^2(x)$ will generally be more alike than those defined by the $v_i^1(x)$. Then, the process can be repeated, in successively anterior original positions $OP_1$, $OP_2$,… $OP_m$,… with the recurrence relation

$$v_i^{m+1}(x) = n^{-1}\Sigma_j H_{ij} \circ v_j^m(x) \qquad (2)$$

for each *x*, all *i*, and all integers *m*. Then, for each *m*, the orderings of the *x* defined by the functions $v_i^{m+1}(x)$ for all *i* are generally more alike than those defined by the functions $v_j^m(x)$, and the Pareto set generally shrinks from one step to the next, in the same sense and for the same reason as above, presented in general form in the appendix. When $m \to \infty$, a full convergence of these individual orderings towards the same ordering represented by the ordinal utility $U(x)$ means that $u_i^m(x) \to u_i^\infty(x)$, $v_i^m(x) \to v_i^\infty(x)$ for all $i$ and $x$, and there are $n$ increasing $\Re \to \Re$ functions $\varphi_i$ such that $u_i^\infty(x) = g_i^{-1} \circ v_i^\infty(x) = \varphi_i \circ U(x)$ with $\varphi_i = h_{ij} \circ \varphi_j$ for all $i$ and $j$.[4] Such a solution satisfies the $n$ limit equations

$$v_i^\infty(x) = n^{-1}\Sigma_j H_{ij} \circ v_j^\infty(x) = n^{-1}\Sigma_j g_i \circ h_{ij} \circ g_j^{-1} \circ v_j^\infty(x)$$

or

$$g_i \circ \varphi_i \circ U(x) = n^{-1}\Sigma_j \gamma_{ij} \circ \varphi_j \circ U(x)$$

for each $x$ and all $i$.

The $v_i^m$, $v_i^\infty$ and such a reached $U$ are by construction increasing functions of the $u_i(x)$ (and do not depend otherwise on state $x$). Such a $U$ thus has the form of a classical "social welfare function" $U(x)=W[\{u_i(x)\}]$. It represents the ordering of all individuals in this *infinite regress* or *fully recursive original position*. Hence, the principle of unanimity in this situation demands that it be the social maximand.[5] As a function of the $u_i$, it has neither an additive utilitarian form (which would be meaningless), nor the structure of additive separability that holds for the $v_i^1$ (and $u_i^1 = g_i^{-1}(v_i^1)$). Unanimity requires more integration of the actual individuals' preferences, so to speak. The increasingness of $W$ in the $u_i$ implies that the final solution is Pareto efficient for individuals' actual preferences.

## 5. The case of a fundamental utility

When writing his original position theory, Harsanyi assumes implicitly that all individuals $i$ have the same function $g_i \circ \tilde{u}_i(b_j, x)$ of $(b_j, x)$ for all $j$ and possible $x$, that is, the same VNM being preferences. This assumes that they have the same being preferences (ordinal) and the same risk-aversion – specifically the same cardinal function $g_i$ of the same specification $\tilde{u}_i$ of the common being preferences. Then, functions $v_i^1$ are the same. These identities are not the case, however. Nevertheless, social choice problems are commonly defined for more restricted populations (set I of individuals $i$) and questions (set $X$ of states x among which to

---

[4] For each $x$, the $n$-vectors $v^m = \{v_i^m\}$ are defined in a compact space from the definition of the $\tilde{u}_i$.

[5] This principle says that if everybody agrees, this opinion should be followed. It is the basis of the theories of the original position of both Harsanyi and Rawls.

choose). Then, similarities in these elements may occur. The first structure concerned has to be that of being (ordinal) preferences, since risk-aversion consists in the appropriate cardinal specification of these preferences. The presence of such properties depends on the case and on the relevant meaning of preference orderings and utilities. Then, in a number of cases, the being preferences are the same, that is, for each state $x \in X$, all individuals have the same preference ordering about being the various individuals. For instance $x$ may simply denote the distribution of incomes and the individuals may be concerned by it just because they prefer to have a higher income.[6] Other aspects of the social states or situations studied may lead to the same structure. In other cases, the $x$ may have a much larger meaning, but there is a conception of "living a better life" in the culture of the society in question, shared by all its members and common to them, which constitutes the being preference ordering.[7] Or the comparison may be about happiness, with a meaning for "happier" or "no less happy than" (including across individuals) for the problem under consideration (sets of states $X$ and individuals $I$), a relation which may constitute the ordering in question, and with a conception of happiness as having a certain objectivity (although possibly depending in particular on mental characteristics).

Hence, in important cases, the individuals agree about the ranking of the desirability of being the various individuals.[8] This common ordering of the pairs $(b_i, x)$ has been called the fundamental preference ordering. When it is representable by an ordinal function, this is the fundamental utility. Let $u(b_i, x)$ denote a specification of this function. Then, for given $i$, $\hat{u}_i(x) = u(b_i, x)$ is a specification of individual $i$'s ordinal utility function for comparing states $x$. The other specifications of the ordinal fundamental utility are $\varphi \circ u(b_i, x) = \varphi \circ \hat{u}_i(x)$, where $\varphi$ is any increasing function. Hence, the functions $\hat{u}_i(x)$ can be replaced by any functions $\varphi \circ \hat{u}_i(x)$ with the same function $\varphi$ for all $i$: that is, they are co-ordinal. However, the other

---

6 This is the topic of an early study in the logical family of original positions by Vickrey (1945). See also Harsanyi (1953).

7 Moreover, in thinly hierarchically ordered societies, the preferences about $b_i$ are commonly obvious to all and shared by all.

8 This is also the case in which thorny psychological problems created by being preferences are absent, such as opposite multiple preferences (preferring to be $i$ rather than $j$ whereas $i$ prefers to be $j$ rather than herself, in particular individuals each preferring to be the other), limits to the actual conception of successive levels of metapreferences (the limit seems to be preferences about preferences about preferences), and weakness of the will (akrasia) about modifying one's preferences.

specifications of individual *i*'s ordinal utility function are $\varphi_i \circ \hat{u}_i(x)$ where $\varphi_i$ is any increasing function (which can depend on individual *i*).

A specification of individual *i*'s cardinal VNM utility is

$$f_i \circ \hat{u}_i = f_i \circ u(b_i, x) = f_i \circ \hat{u}_i(x) = v_i(x),$$

where $f_i$ is an appropriate increasing function (its cardinality says that it can be replaced by any function $a_i f_i + b_i$ where $a_i$ and $b_i$ are constant and $a_i > 0$).

Then, when individual *i* in the original position considers the prospect of becoming any of the individuals with an equal probability $1/n$, she orders the states of the world *x* with her VNM expected utility of the corresponding risk:

$$v_i^1(x) = n^{-1}\Sigma_j f_i \circ u(b_j, x) = n^{-1}\Sigma_j f_i \circ \hat{u}_j(x) = n^{-1}\Sigma_j f_i \circ f_j^{-1} \circ v_j(x).$$

This order is also represented in terms of the fundamental utility levels as

$$\hat{u}_i^1(x) = f_i^{-1} \circ v_i^1(x) = f_i^{-1}[n^{-1}\Sigma_j f_i \circ \hat{u}_j(x)],$$

or, denoting as $M[\{\alpha_i\}, \varphi] = \varphi^{-1}[n^{-1}\Sigma\varphi(\alpha_i)]$ the generalized mean of the *n* numbers $\alpha_i$ with function $\varphi$,

$$\hat{u}_i^1(x) = M[\{\hat{u}_j(x)\}, f_i].$$

Since this case is a subcase of that of the previous section, that in which $h_{ij}=1$ for all *i* and *j*, the discussion of the general case again applies here. There is a multiplicity of evaluations in the original position, one for each individual. And yet the orderings of *x* implied by the $\hat{u}_i^1$ are more alike than those implied by the $\hat{u}_i$, and the Pareto set shrinks (with possible limiting cases, see the appendix). The solution to the problem raised by the obtained multiplicity which is consistent with a use of an original position in the first place consists of considering an original position of the original position, and so on.

Then, $OP_{m+1}$ obtains from $OP_m$ with the evaluation functions

$$\hat{u}_i^{m+1}(x) = M[\{\hat{u}_j^m(x)\}, f_i]$$

in fundamental utility and

$$v_i^{m+1}(x) = n^{-1}\Sigma_j f_i \circ f_j^{-1} \circ v_j^m(x)$$

for the VNM utilities, for all $i$. This constitutes, again, a multiplicity of evaluations, but with implied orderings which are generally more alike, and in general a shrinking of the Pareto set.

If $m \to \infty$, then $\hat{u}_i^m \to \hat{u}_i^\infty$ for all $i$, the $\hat{u}_i^\infty$ satisfy

$$\hat{u}_i^\infty(x) = M[\{\hat{u}_j^\infty(x)\}, f_i]$$

or

$$n\, f_i \circ \hat{u}_i^\infty(x) = \Sigma_j\, f_i \circ \hat{u}_j^\infty(x),$$

for all $i$. These conditions are satisfied if and only if the functions $\hat{u}_i^\infty(x)$ are the same (the levels are the same for each $x$, given that functions $f_i$ are increasing and $n \geq 2$). Thus, we have for all i the same function $\hat{u}_i^\infty(x) = U(x)$. Hence the $\hat{u}_i^\infty(x)$ have the same value for the possible $x$ that maximizes them, i.e. $U(x)$. This condition of an equal level of happiness or satisfaction (in fundamental utility) is "eudemonistic justice"[9]. However, this is for the individuals in the "infinite original position", not for the actual individuals with their actual preferences.

From its construction, $U(x) = W[\{\hat{u}_i(x)\}]$. Function $W$ is an increasing symmetrical function of the $\hat{u}_i$ (at each step, each $\hat{u}_i^{m+1}$ is an increasing symmetrical function of the $\hat{u}_i^m$, and each $\hat{u}_i^1$ is of the $\hat{u}_i$). This increasingness guarantees the Pareto efficiency of the result. The symmetry implies the corresponding impartiality; it is meaningful only because of the existence of a fundamental utility. If all individuals are very risk-averse, for all $i$ $\hat{u}_i^1(x) = \hat{u}_i^\infty(x) = U(x) = \text{Min}_j\, \hat{u}_j(x)$, which is eudemonistic "practical justice"[10]. Note that the direct equality of the $\hat{u}_i(x)$ may have to violate Pareto efficiency, or may not be possible. This was one reason for resorting to practical justice. However, this solution was too extreme for a general solution. Finally, if all functions $f_i$ were the same (cardinally, that is, up to an affine increasing function) and were function $f$ (that is, $f_i = a_i f + b_i$ with constant $a_i$ and $b_i$ and $a_i > 0$, for all $i$), then, for all $i$ and $m \geq 1$, $U(x) = \hat{u}_i^m(x) = \hat{u}_i^1(x) = M[\{\hat{u}_j(x)\}, f]$ and a maximand can be $nf \circ U(x) = \Sigma f \circ \hat{u}_i(x) = \Sigma v_i(x)$, calling $f(\hat{u}_i) = v_i$. This was the form intended by Harsanyi.

---

[9] See Kolm 1971.

[10] Id. "Practical justice" was more generally defined as the leximin in the $u_i(x)$.

It requires both a fundamental utility and identical preferences concerning risk with respect to it.

## 6. Related other solutions

### *6.1. Agreement in the original position*

An alternative solution consists of agreements of the individuals about the choice of *x*. A number of theories study it (bargaining). When it is given a moral value, this is because of the freedom manifested by the free agreement. However, this implies the moral endorsement of all the elements that determine the outcome (threat point, bargaining power, time preference, etc.). However, still another solution consists of using agreement for solving the problem of the plurality of individual views in the original position only. The agreement, then, is hypothetical, notional, and the individuals' utilities are their $u_i^1$ or $v_i^1$ in the original position. These evaluations agree more than the $u_i$ or $v_i$. The original position theory solves part of the problem. Giving a moral value to a hypothetical agreement is one of the most classical social ethical method, since this is, by definition, a social contract.[11]

### *6.2. Comparabilities in economics*

Mentioning or writing preferences about "being" of some sort have by now a notable history in economics. The important point, however, is not writing but *meaning*. There are a number of cases. Some are just mention and others are formal writing. The evaluation can be an ordering, an ordinal utility, or a cardinal utility. In the latter case, this is either a VNM utility or just a cardinal utility (often thought to also necessarily be the former). "Being" is sometimes restricted to a preference ordering. The evaluation is either interpersonally comparable or it is not. Harsanyi (1955, 1976, 1977) considers a comparable VNM cardinal utility as the universal case, a problematic assumption. Tinbergen's (1957) discussion implies comparability for "equal happiness" and needs no more than ordinalism. Arrow's (1963)

[11] This is how Rawls introduces the theory of the original position: as the "state of nature" of the classical theory of the social contract. However, since the individuals he envisions in the original position prefer the same social state (defined by his "principles of justice"), they agree a priori and there is no point to add another agreement by a contract (except, perhaps, as a mutual promise to implement these principles and the resulting state, which binds morally the actual individuals that the "original" ones becomes when the veil of ignorance is lifted).

mention of "extended sympathy" is ordinal non-comparable. Kolm presents ordinal non-comparability and comparability (1966) and an extensive use of ordinal comparability (1971). Pattanaik's (1968, 1971) example argues for a comparable specification in certainty and non-comparable VNM utility for uncertainty; his example is a case of the original position used in section 5, although without solution to the problem of the multiplicity of individual evaluations. The ordinal comparability of fundamental preferences has then had a number of uses (Hammond (1976), Arrow (1977), Becker and Stigler (1977), and others)[12]. Individuals' preferences about both consumption and an individual preference ordering are considered by Sen (1970), Suzumura (1983), Mongin and d'Aspremont (1998), and Mongin (2001) – the latter for VNM cardinal utilities. One should finally note the case of utilitarianism, which requires cardinal individual utilities defined up to a common multiplicative factor (co-multiplicative cardinality).

### *6.3. Limitation to theories of the original position*

The ethically delicate part of a social choice is the justice of the outcome, notably concerning distributive justice. A theory of the original position amounts to the assimilation of such a choice, and in particular the choice of justice, to a self-interested choice in uncertainty. This assimilation has limits. One may bet all one's wealth on a single horse in the anticipation of the possible great pleasure to be millionaire. Then, may one give all the wealth of a society to a single person so that there exists the experience – assumed pleasurable – of being a millionaire? The problem is one of responsibility. A person with a sane mind is responsible for the risks she takes concerning herself. By contrast, a choice of justice is accountable towards all people, society, and morals. The inequality-aversion of a choice of justice cannot be the formal equivalent of some individual risk-aversions that are considered no reason to

[12] The normative part of this literature frequently attributes the maximin in fundamental utility of "practical justice" to Rawls, whereas the first tenet of Rawls is a rejection of a concept of utility and he explicitly rejected this attribution: "to interpret the difference principle [his principle of a maximin in an index of "primary good"] as a maximin in utility is a serious mistake from a philosophical point of view" (Rawls, 1982). Rawls (1982), however, favourably discusses the concept of fundamental preferences or utility. The point is that Rawls's principles of justice are only general principles for overall distributions at national levels. The scope of application of the maximin or leximin of "practical justice" was not specified. They are justified in situations in which the lowest utilities mean serious suffering which can be sufficiently remedied by the policy (it is then not seriously ambiguous that these people can be considered the least happy).

interfere with the individual choice.[13] It has a priori to be higher. [14] Moreover, the individual choice leads to consider, as compared material, "utility" or satisfaction. By contrast the choice of justice may compare means of satisfaction, such as incomes or more generally Rawls's primary goods for overall distributive justice (Rawls' "social justice", "macrojustice"). The conclusion is that the equivalence between individual risk and social justice cannot be applied without a specific discussion which a priori depends on the case. A criterion may be that it is generally considered valid by the people concerned.[15]

## 7. Appendix. Homogeneization and convergence of individual preference orderings

Let us make precise the noted relations concerning the $u_i^m$ (or, equivalently, the $v_i^m$). Denote $u_i = u_i^\circ$; $m$ the successive integers $m$=0,1,2,…; and $u^m = \{u_i^m\}$ the $n$-vector of the $u_i^m$ for all $i$. Denote as usual, for the $n$-vectors $y=\{y_i\}$ and $z=\{z_i\}$, $y=z$ as $y_i=z_i$ for all $i$; $y>z$ as $y_i>z_i$ for all $i$; $y$ $y$ as $y_i \geq z_i$ for all $i$; and $y \geq z$ as $y_i \geq z_i$ for all $i$ and $y_i>z_i$ for at least one $i$. From equations (1) and (2) and $u_i^m = g_i^{-1} \circ v_i^m$ where $g_i$ is increasing for all $i$ and $m$, then, for all $i$ and $m$, $u_i^{m+1}$ is an increasing function of $u_j^m$ for all $j$. Denote as $x$ and $x'$ two social states. Therefore,

$$u^m(x) = u^m(x') \Rightarrow u^{m+1}(x) = u^{m+1}(x'),$$
$$u^m(x) > u^m(x') \Rightarrow u^{m+1}(x) > u^{m+1}(x'),$$
$$u^m(x) \quad u^m(x') \Rightarrow u^{m+1}(x) \quad u^{m+1}(x'),$$

and

$$u^m(x) \geq u^m(x') \Rightarrow u^{m+1}(x) > u^{m+1}(x') \Rightarrow u^{m+1}(x) \geq u^{m+1}(x'). \tag{3}$$

Hence, these four types of unanimous preferences between two social states are maintained from each stage to the next. The set of pairs of states related by one of these unanimous

---

[13] Structural differences between inequality-aversion and risk-aversion are also suggested by enquiries, questionnaires and experiments (Amiel and Cowell (1999), Kolm (2001)).

[14] For instance, in the case of a fundamental utility, one may use a kind of maximin risk-aversion by taking, as social maximand, $\Sigma f \circ \hat{u}_i(x)$ where the cardinal VNM function $f$ is such that, for each level $\hat{u}$, $-f''/f' = \max_i - f_i''/f_i'$.

[15] An alternative to the original position is the theory of *moral time sharing*, i.e., an individual considers that she is all the actual individuals successively in time, perhaps recursively, for the same duration (or with adjustment for discounting depending on specification). This can also be related to a theory of the multiple self, with dated selves and notional individuals who have the actual individuals as their various selves. These theories raise issues similar to that of the original position.

preferences looses no element. Since it changes in general, this implies that it expands. In particular, the set of states that are unanimously indifferent, strictly preferred, preferred or indifferent, or this relation plus at least one strict preference, to a given one, or that a given one equals or dominates in any of these senses, looses no element and generally expands. In this sense the preference orderings become more alike.

Denote as $\Pi$ the set of possible states *x* and as $P^m \subseteq \Pi$ the set of Pareto-efficient states with preferences of order *m*. Then, if $x' \in P^{m+1}$, from the definition of Pareto efficiency there is no $x \in \Pi$ such that $u^{m+1}(x) \geq u^{m+1}(x')$. Hence, from relation (3) there is no $x \in \Pi$ such that $u^m(x) \geq u^m(x')$. That is, $x' \in P^m$. Therefore, $P^{m+1} \subseteq P^m$. Since, in general, $P^{m+1} \neq P^m$, this implies $P^{m+1} \subset P^m$. Generally, $P^\circ \supseteq P^1 \supseteq \ldots P^m \supseteq P^{m+1} \supseteq \ldots$[16]

**References**


**Amiel, Y. and F. Cowell.** 1999. *Thinking about Inequality: Personal Judgment and Income Distribution*. Cambridge: Cambridge University Press.

**Archibald, G.C., and D. Donaldson.** 1979. "Notes on economic inequality." *Journal of Public Economics*, 12: 205-14.

**Arrow, K.** 1977. "Extended sympathy and the possibility of social choice." *American Economic Review. Papers and Proceedings*, 67: 219-229.

**Becker, G. and G. Stigler.** 1977. "De gustibus non est disputandum." *American Economic Review*, 67: 76-90.

**Edgeworth, Y.** 1888. *Mathematical Psychics*.

**Hammond, P.** 1976. "Equity, Arrow's conditions and Rawls' difference principle." *Econometrica*, 44: 793-804.

**Harsanyi, J.** 1953. "Cardinal utility in welfare economics and in the theory of risk-taking." *Journal of Political Economy*, 61: 434-35.

**Harsanyi, J.** 1955. "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility." *Journal of Political Economy*, 63: 309-21.


[16] The same logic for two levels and for the question of altruism is considered by Edgeworth (1888) in a simple case and by Winter (1969) and Archibald and Donaldson (1979). It is analyzed for the infinite recursive chain by Kolm (1984, 2000) in general and in application to the influence between individuals' preferences in a dialog.

**Harsanyi, J.** 1976. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: P. Reidl.

**Hicks, J.** 1959. *Essays in World Economy.* Oxford: Basil Blackwell. Preface reprinted as "A Manifesto". In *Wealth and Welfare*. Oxford: Basil Blackwell, 1981, pp. 135-41.

**Kolm, S.-Ch.** 1966. "The optimal production of social justice." International Economic Association conference on Public Economics, proceedings, Biarritz. H. Guitton and J. Margolis, eds. *Economie Publique,* Paris: CNRS, 1968, 109-77, and *Public Economics*, London, Macmillan, 1969, 145-200.

**Kolm, S.-Ch.** 1971. *Justice et Equité*. Paris: Cepremap (Reprinted 1972, Paris: CNRS). English translation. 1998. *Justice and Equity*. Cambridge MA: MIT Press.

**Kolm, S.-Ch.** 2000. "Introduction." In *The Economics of Reciprocity, Giving and Altruism*, ed. L.-A. Gérard-Varet et al. London: Macmillan.

**Kolm, S.-Ch.** 2001. "Unequal Thoughts." *Journal of Economics,* 73(1): 96-106.

**Mongin, P.** 2001. "The impartial observer theorem of social ethics." *Economics and Philosophy*, 17(2).

**Mongin, P. and C. d'Aspremont.** 1998. "Utility theory in ethics." In *Handbook of Utility Theory*, Vol.1, ed. S. Barberá, P. Hammond and C. Seidl, 342-73. Boston: Kluwer Academic Press.

**Pattanaik, P.** 1968. "Risk, impersonality, and the social welfare function." *Journal of Political Economy*, 76: 1152-69.

**Pattanaik, P.** 1971. *Voting and Collective Choice*. Cambridge: Cambridge University Press.

**Rawls, J.** 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

**Rawls, J.** 1982. "Social unity and primary goods." In *Utilitarianism and beyond*, ed. A. Sen and B. Williams. Cambridge: Cambridge University Press.

**Sen, A.** 1970. *Collective Choice and Social Welfare.* Amsterdam: North Holland.

**Suzumura, K.** 1983. *Rational Choice, Collective Decisions and Social Welfare*. Cambridge: Cambridge University Press.

**Tinbergen, J.** 1957. "Welfare economics and income distribution." *American Economic Review*, 47: 490-503.

**Vickrey, W.** 1945. "Measuring marginal utility by reaction to risk." *Econometrica*, 13: 319-33.

**Winter, S.** 1969. "A simple model of the second optimality theorem of welfare economics." *Journal of Economic Theory*, 1: 99-103.