

May 2001

**VOX POPULI, VOX DEI:
ENDOGENOUS SOCIAL CHOICE AND THE RATIONAL ORIGINAL POSITION**

Serge-Christophe KOLM

Abstract

Respectful or endogenous social choice is social choice wanted by the members of the society. Unanimity is hampered by self-centeredness which, however, should be laundered away for revealing individuals' conceptions of justice which should be impartial. The remaining much smaller differences in individuals' views are resolved in various ways. The conceptual problem of each individual jointly caring for all others as she does for herself is notably faced by the theories of the original position, whose problem raised by the necessary multiplicity of individuals' original evaluations is consistently faced by a regress of original positions, and can also be faced by original aggregations or a genuine original social contract. However, the differences between a choice of justice and an individual choice in uncertainty lead to other solutions which consider justice as the notional general extension of other close social relations between persons.

Keywords: Social choice, original position, justice, fundamental utility.

1. Presentation¹

1.1 The essence: unanimous impartiality

Endogenous social choice derives ought from is. Since I do not know what is good for society (who dares say she knows?), and since, if I thought I knew, there would probably be no reason to apply my view rather than others', I can only look for an answer in the observation of society itself. In fact, there exists no possible alternative for the society made of everybody (including you and me). Of course, society members often seem to disagree. But most of their disagreement stem from their opposed self-interest. This is not their moral, impartial view. And if some seem to have no such impartial view, it is possible to rationally infer one from their actual views. These impartial views will be much closer to one another than the others, tainted by self-interest. And there are ways of seeing them agree in the end, such as through impartial views of impartial views, agreements among these views, or mutual information about causes of moral views and feelings that accompany them. This paper consists of making these points explicit.

1.2 Summary: Intent, problems, and solutions

A summary of the main issues, later to be analyzed, is proposed in this section.

An endogenous solution to social choice derives social rules from the views of society's members. In a sense, there exists no other alternative for the all-inclusive society comprehending everybody. For smaller societies, external observers have a priori no reason to sufficiently and correctly care, and at any rate should not intrude if democracy and autonomy are to be respected. Then, unanimity and consensus among society's members have to be endorsed. However, self-interest and self-centeredness make individuals' views opposed to one another in questions of distribution. But they should be discarded in individuals' explicit or implicit conceptions of justice whose general property is impartiality. Hence, the first and an important step consists of the notional impartialization of individuals' views. Impartializations which respects the most people's views consist of assuming the extension to society as a whole of individuals' ways of judging and sharing among people who have a social relation of the same type with them (family, neighbours, common members of society, etc.).

This solution is probably the most moral the higher the concern for others implied by this social relation of reference, and hence the closest this relation. A limiting case consists of caring for all others as for oneself. This solution, however, raises the conceptual problems of considering being several individuals jointly, of assimilating with each of these different persons and situations, and of comparing them. Among the possible solutions, such as moral time-sharing (successively being the various individuals) or theories of the multiple self, the theories of the "original position" are the most developed. Then, the individuals in an original position should envision they could become any of the actual individuals, and any larger uncertainty is superfluous and hence arbitrary (this is the "thin veil of ignorance"). For this purpose of building an ethical theory, impartiality will justify equal probabilities. However, an individual in the original position cannot but evaluate with her own preferences

¹ I have much benefitted from Philippe Mongin's criticism and from his impressive extended exegesis of Harsanyi's utilitarian proposals, and from comments from Maurice Salles, Marc Fleurbaey, Henri Nzitat, Antoinette Bujard, and other participants at a seminar at the University of Caen. I alone am responsible for any error.

about risk. Moreover, individuals can a priori have different preferences about becoming various individuals. These constitute two reasons why individuals in the original position have different evaluations. The theory, therefore, fails to provide a unique, consensual evaluation in the original position. An important progress has nevertheless been made in that individuals' preference in the original position are much more similar than the actual individuals' preferences are.

The concept of preferences about "being" various possible individuals raises difficulties, notably logical ones, which, however, will not interfere with the proposed solutions. Moreover, some of these conceptual difficulties, and the second of the two reasons for the differences in individuals' evaluations in the original position, vanish when the individuals, in each state of the world, agree about the order of desirability of being the various individuals. This case of a "fundamental utility" avoids a type of inconsistency in the structure of preferences about being alternative individuals, and it can result from a common social conception of happiness or of the good life. It permits specifying the solutions and proposing specific solutions, but the a priori multiplicity of evaluations in the original position remains for the former reason (different individual preferences toward risk).

Three types of solution to the problem created by the multiplicity of individual evaluations in the original position will be proposed. One solution, the "regress", is in fact demanded by consistency. Indeed, since the theory of the original position intended to solve the problem of the variety in individuals' views in the first place, and a progress has been made, both internal consistency and the reasons for resorting to an original position in the first place require solving the obtained problem of multiplicity of individuals' views in the original position by an original position of the original position. This, again, will leave one evaluation by individual, but they will be more similar. And so on it goes. The situation converges to consensus, with increasing similarity of individuals' evaluations, and a shrinking of the Pareto set (the set of Pareto-efficient states).

The two other solutions to this problem of multiplicity consist of considering, in the original position, an aggregate individual and a genuine social contract². The former yields specific solutions in the case of a fundamental utility, as a result of the structure of the functions involved.

The theory of the original position happens to have a central place in contemporary social ethical theory, since it was used by Rawls, in a milieu dominated by utilitarianism, for opposing it and replacing it with the classical basic rights and a traditional ideal of equality in incomes and other primary goods, while the same theory, as used by Harsanyi, is a main, the main, or the only, justification of a utilitarian form for most present-day utilitarians. Hence, showing the problems of the theory of the original solution, which make it differ from these two contending views, and the various possible solutions to these problems, seems important for contemporary social ethics. The problem of multiplicity (of views in the original position) can be solved as just noted. However, the following problem in the possibility of reducing justice to individual risk leads to resorting to different theories of endogenous social choice, with concerns for others which fall short of identification.

² This is some agreed upon compromise among the individuals' views. Rawls considers a social contract in an original position, but the individuals, there, a priori want the same thing and hence need no compromise, as a result of Rawls's particular form of an original position and assumptions about it.

Indeed, the choice of justice among individuals is not a priori isomorphic to an individual's choice in uncertainty. An individual taking chances for himself has the right to accept more unfavorable possible outcomes than has a choice of justice when these outcomes mean the situations of different individuals. Moreover, justice may leave some items to individuals' private spheres and choices. Similar problems affect moral time-sharing.

Then, an endogenous solution to social choice has to resort to a reference to close social relations which are not identification with others. However, the corresponding individuals' impartialized views may differ, in spite of the homogeneizing effect of social norms. The solution then consists of assuming full – and hence equal – information of the individuals, in an extensive sense. This includes fully knowing others' "formative information", that is, the information about facts, values and life experiences that shape moral views. It also includes "empathetic information", that is, knowing not only others' moral views but also how it feels to fully have the corresponding sentiments. This information and the resulting homogeneization of individuals' views can be promoted by the dialog of "discourse ethics".

1.3 Overview

This text follows this general summary.

The general philosophy of the endogenous solution to social choice is outlined in section 2. Endogeneity and endorsing informed unanimity or consensus are necessary or practically so (2.1). However, justice requires impartiality, and hence the crucial concept is the notional respectful impartialization of members' views (2.2). And yet, the resulting individual views, though much closer to one another, are still different, and this "problem of multiplicity" must be solved, notably through a regress of impartialization; however, there remains the problem of choosing the social relation or distance of reference for the impartialization (2.3). The limiting case of full concern for others – jointly caring for all *alters* as for *ego* – requires specific conceptual solutions such as an original position or moral time-sharing, with solutions to the problem of the remaining multiplicity; possible misgivings about these solutions then lead to the possible choice of other impartializations – such as extending to society relations among kin (2.4).

The structural issues of the theories of the original position are shown in section 3. The "thin veil of ignorance" is to be favored (3.1). Having preferences about "being" someone raises problems, and the two reasons for the individual specificity of the evaluations in the original position are noted, while the present social ethical concern justifies equal probabilities (3.2). These evaluations constitute an important progress toward consensus, and the solutions to the problem of their multiplicity consist of making them identical by regress, aggregation or a genuine social contract (3-3). The case of fundamental preferences and utility is discussed in section 4. The theory of the "rational original position", which solves the multiplicity problem by a consistent and convergent infinite regress, is analyzed in section 5, in distinguishing the general case (5.1) and the case where a fundamental utility can be assumed (5.2), while making precise the issue of the convergence of individuals' evaluations (5.3). The theory of the aggregate original evaluator and its necessary results constitute another way of solving the problem of multiplicity (section 6), as is an original social contract (section 7).

However, the intrinsic differences between a choice of justice and an individual choice in uncertainty, presented in section 8, lead one to consider other solutions either still with full concern such as moral time-sharing (section 9), or with other social relations of reference (section 10), with proposed solutions for the still present problem of multiplicity through an extensive conception of the role of information in the formation and transformation of moral views.

2. The endogenous solutions to social choice

2.1 Endogeneity and unanimity

2.1.1 Reason and necessity of the endogenous solution to social choice

Where can the social ethical rules of a society be derived from? One possible answer can be, or can begin with: from the views of the members of the society. The first reason for this answer is epistemic: where else can we look for the answer? A reason why one state of society would be better than another can only occur in some person's mind. If the considered society is the set of all persons – the grand society – no other mind exists, there is no external observer (even when some society's members see their own view this way). If the considered society is smaller, there can exist foreign, external observers. But, first, why would they care, at least with sufficient consideration of the issues, problems, and situations? If there is some external effect from the state of this society to some other person, this should probably be solved in including this person within the considered society. More importantly, resorting to some foreigner's view would a priori violate a condition of democracy, autonomy, independence, self-rule or home rule, collective freedom, respect of political basic rights, non-dictatorship, and non-paternalism for this society. And if a foreign observer wants to respect the views of this society's members, she would again base her consideration on these people's opinions. In other words, we are looking for an *endogenous* solution to the problem of social choice.

By an endogenous solution, we mean a fully endogenous one. Most imaginable solutions have some endogeneity. If not, this would probably not be social choice at all. For instance, majority voting uses members' preferences, and a classical social welfare function aggregates people's welfares. But we also want to derive from the society, as the case may be, which voting rule and why voting, or which welfare function and why such a solution. Hence, we shall want to find out what people relevantly really want about the problem considered. However, this may not be straightforwardly observed because people's desires may mix several motives – for instance self-interest and an explicit or implicit sense of justice –; their interests may prevail over their moral or social views in their expression or observable behaviour; they may fully lack the required social ethical views (as when sharing among purely selfish people); or they may disagree on moral grounds as they do about their self-interest. Hence, social ethical views may have to be disentangled from others, or fully inferred or justifiably constructed from other aspects of people's psychology, and we may be left with a plurality of views.

2.1.2 The philosophy of the solution: moral consensus from informed impartiality

However, even without knowing society's or its members' views, we know a priori important things about these views, namely the intrinsic, necessary and definitional properties of views concerning moral, social ethics, and notably justice. These properties concern two aspects:

impartiality for justice, and information. For a reason of rationality, the relevant ethical views cannot be based on ignorance, confusion or mistaken beliefs. They should be fully informed about the relevant issues and facts, at least as a notional, hypothetical and conceptual property. Moreover, the aspects of social ethics that concern the good of individuals (or other social entities) is justice (“respectful justice” if the individuals agree that the considered items are good for them)³. And the basic property of justice is impartiality. Impartiality opposes exclusive self-interest or self-centeredness and will shortly be further analyzed⁴.

A consequence of these necessary properties of individuals’ social ethical views is that these views will turn out to be much closer to one another than individuals’ actual views much influenced by self-interests, self-centeredness and different information of various types. That is, the scope of consensus, agreement, or coincidence of individuals’ judgments, will greatly expand; the domains of disagreement and opposition will shrink importantly. Impartiality clearly has this effect for problems of distribution where individuals’ interests are opposed by nature. It is not so clear a priori that information has this effect (mistaken beliefs may favor agreement), but this will turn out to be the case for the full information we shall consider. Now, we shall shortly see that endorsing unanimous or consensual judgments is implied by the very reason for an endogenous solution. This will provide the way to this solution. In the end, the full solution will result from sufficiently full impartiality and information which make individuals’ views coincide, as we shall see. Hence, a basic exercise will be to find the rational and moral unanimity implied by different and opposed views – for instance the completely opposed self-interests when sharing scarce resources.

2.1.3. The moral of unanimity or consensus.

Indeed, if all society’s members agree about a solution, an evaluation or a preference, an endogenous moral solution probably has to endorse this view if it is sufficiently informed. The first reason is epistemic. If the considered society includes everybody – the grand society –, then this opinion is the sole existing one about this topic. No other view exists. In particular, you and me, who are within this society, hold this judgment – as Pascal said, we are on board. No other view can be heard or read as a sincere expression.

If, on the other hand, we consider only a smaller society, then the condition of endogeneity just discussed, saying that society’s rules depend on its members’ views only and on nobody else’s, yields the same conclusion for the same reasons (except, possibly, the reference to our own opinion). Of course, unanimity precludes that some society’s members be harmed by others in choices which implement the considered evaluation. We would just have to ascertain that this unanimous view does not damage anybody or any valuable thing outside this society.

2.1.4. Qualifications

³ The basic concepts concerning justice are presented in Kolm 1996, chap.1.

⁴ Conceptually, impartiality may be a consequence of the notional full information in the sense that one may be unable not to be impartial if one knows others’ situation, feelings, etc. as much as one’s own. This is of course a purely hypothetical situation, and the reduction of impartiality to an issue of information is a speculation which has some interest in relating norms to the single issue of information, but is not necessary – and then impartiality is taken as a primary necessary and definitional property of a view of justice.

However, this respect of other persons can be taken care of by the requirement that the considered society includes all persons so concerned, so as to internalize such externalities. For instance, the views of future generations may have to be considered for their education or the maintenance of the environment (the relevant aspects of their views should then be reasonably guessed at). More generally, the principles of unanimity and more generally of endogeneity may be submitted to conditions or to restrictions of domain. We may require that the society respects values that transcend its members' possible actual preferences, such as cultural heritage, environment or children's education. Moreover, the value of social autonomy derives, in the end, from the autonomy of the judgments of society's members (in the sense of Rousseau and Kant), and the latter can be violated by mischievous leadership, vicious tradition, or irrational contagious collective enthusiasm. However, we have subjected the value of endogeneity – and hence that of unanimity – to the condition that the individuals be sufficiently informed. Now, the concept of information is amenable to a large extension – as we shall see. In particular, the use of reason leading to mental autonomy can be seen as increasing information. And information may also erase insufficient knowledge of the value of cultural heritage and environment. At any rate, the application of principles can be submitted to corresponding restrictions of domain. Moreover, our principal concern will be the “simple” question of the distribution of scarce resources.

2.2 Impartiality and impartialization

2.2.1 Impartiality

The question of the allocation of resources, however, is precisely one where people have a major reason not to agree, since this distribution makes their interests opposed to one another. This is what makes an endogenous and unanimous solution to this problem notable, powerful, and possibly subtle (or tricky) to establish. This problem is that of distributive justice. The characteristic property of principles of justice is their impartiality. Impartiality is the opposite of self-centeredness, that is, claiming something for someone because of her particular relation to the claimer, such as being the claimer herself, a relative or friend of hers, a member of a social group of any kind to which the claimer also belongs (same region, nation, language, race, age, sex, profession, club, and so on), or someone the claimer likes and who has characteristics she likes. Exclusive self-interest is a particular case of self-centeredness, and a pure breach of impartiality.

However, many allocative criteria satisfy the principle of impartiality. For instance, principles of the type “to each according to...” can refer to a priori many things while claiming impartiality. Impartiality only consists of seeing everyone “on the same footing” in a sense, but this can mean many possible things. Formally, impartiality will be a property of *permutability* of the aspects of individuals' situations chosen as relevant in a rule (these aspects jointly include the evaluated items and the characteristics to which they are related in the rule). For instance, an impartial evaluation function will be *symmetrical* in these aspects (this means invariance under such permutations). But this is only one structural aspect of these rules or evaluations. There remains to choose the relevant aspects of individuals' situations, as regards both the evaluated items and individuals' characteristics to which they should be related, the formal structure of this relation, and the structural properties of the rule or evaluation function other than its symmetry⁵.

⁵ A very common mistaken view consists of believing that impartiality suffices to solve the question of optimality and justice (this affects many notable authors, Adam Smith to begin with). Now, for

2.2.2 *The two sources of views of justice*

Individuals are concerned with what happens to themselves. This is their self-interested judgment. They usually also are particularly concerned with their family, their friends, and, more or less, with other people having something in common with them or whom they like. Along with their self-interest, this constitutes their self-centeredness. In addition, most people have views about justice. I do not mean here their possible “phariseanism”, that is, defending one’s interests or those of the people one likes with ad hoc arguments referring to moral and notably justice, whether or not one comes to believe or half-believe in these reasons. Rather, I refer to a genuinely moral view, a sense of justice with notably a property of impartiality. This is the judgment of conscience, of what Adam Smith calls the “impartial spectator” within the individual in his *Theory of Moral Sentiments*. This view is the one we want to see have an influence on facts of basic distribution. It is in particular the view we need to know for determining the endogenously just solution.

However, the influence of this view on facts meets various important difficulties. Most people, most of the time, support much more their interest than their moral view about justice. This is explainable by the fact that justice is a collective concern. In particular, people defend their interests or those of people they favour because other people do the same, in order to keep a balance of forces or arguments. However, people also generally support their interest when their action or expression has no actual effect, such as with voting in large elections or answering in large polls. Moreover, people’s conception of justice often reflects spurious social influences (that of the family or its opposite). The moral arguments people present or consider aim more often at justifying or rationalizing already adopted opinions than at building such an opinion. And, not uncommonly, expressed moral arguments in fact only defend interests (phariseanism). Moreover, people often adapt their moral views to their interest, in defining the characteristics deemed relevant (while keeping an impartial form), for avoiding the mental pains of cognitive dissonance, split personalities, bad conscience, shame, or self-indictment. In fact, individuals are frequently unclear to themselves about the moral validity or status of such views of theirs which happen to support their interests.

Hence, though people’s explicit moral views will remain an important source of information, notably when they result from dialog or deliberation, the most important way of obtaining people’s “true” views about justice will consist of inferring or constructing them from their other judgments and behaviour. For instance, an individual’s judgment when she compares or shares among people equally dear to her (for instance her children, or on the contrary distant members of the society) can be notionally extended to the whole of society including herself (the “extension” method). Or the individual may be assumed to be all society’s individuals successively in time (“moral time-sharing”) or as uncertain possibilities (an “original position”).

The essence of an endogenous solution is to keep as much of the individuals’ actual preferences as possible. However, a view of justice has to have the formal structure of impartiality. Hence, the endogenous solution to the problem of justice begins with notionally transforming, as little as possible, individuals’ preferences about society, for making them impartial. The actual material so transformed can be self-interest (as with moral time-sharing

example, impartiality can be satisfied both with equal incomes and with the full self-ownership of classical process-liberalism.

or an original position), or social views which can be self-centered (as with “extension”). This notional impartialization of people’s views derives, from their actual views, judgments they would have were they only impartial. It launders partiality and self-centered biases away from people’s views. But it otherwise changes these views as little as possible, for keeping with endogeneity. It maximally respects these views, and only provides them with the service of making them moral from society’s standpoint. Hence, the endogenous solution to social choice fully respects individuals’ moral views, abides by them, and implements them, be they explicit opinions or impartial evaluations implicit in actual views and implied by them.

2.2.3 Respectful impartialization by extension

Actual people evaluate actual people’s situations. This judgment generally depends on the social relation between the evaluator and the evaluated because of the former’s self-centeredness. For instance, the evaluator is bound to judge these situations differently because they are those of complete foreigners, common members of her society, neighbours, first degree cousins that she sees once a year, brothers, children, or the evaluator herself. These causes of differences would not exist in an impartial judgment. Hence, with respect to this issue, a notional impartialization of an individual’s view, which keeps as much as possible of the features of this view, consists of imagining that all the evaluated situations are those of individuals in one same relation with the evaluator (all foreigners, common society’s members, cousins, brothers, and so on). The definition of this relation can include any cause of the individual’s behaviour toward others (general or particular liking, sense of duty, norm, etc.). In the extension and impartialization, the evaluator’s own situation is included as one among these evaluated situations. Then, the impartialized individual evaluation will duplicate the actual individual’s view of the situation of individuals with whom she actually is in this relation. We would, for instance, extend to the whole society this individual’s ways of sharing among her brothers, children, cousins, or neighbours, or among common members of the considered society, or perfect foreigners. This relation is the *social relation of reference for the impartialization*. The individuals’ impartialized views then rest on an actual feature of individuals’ judgments.

The pure logic of the foregoing impartialization of individuals’ evaluations can be presented as follows. Individuals will be represented by indexes i, j , etc. Let s_i denote the “situation” of individual i , that is, the set of all possible facts or parameters concerning this individual and which may be found relevant for social ethics and notably justice. Some of these parameters are the object of the evaluation and will a priori be directly or indirectly affected by the policy, and others can denote individual i ’s characteristics deemed relevant for this evaluation. The former category can for instance include the individual’s income, wealth, general or specific consumption, welfare, index of level of happiness or a given type, social position and various aspects of it, and so on (and the second category can for instance include need, capacities for enjoyment, pain, effort, labor, merit, natural rights, and so on). Let r_j^i denote an index (of any nature) of the type of relation of individual j to individual i (for instance, individual j is a brother, neighbour, child, fellow citizen, complete foreigner, and so on, for individual i). Then, individual i ’s evaluation of the state of society is assumed to be described by the level of the ordinal function $u^i(\{s_j, r_j^i\}_j)$, depending on all the s_j and r_j^i and symmetrical in the pairs (s_j, r_j^i) for each j . One particular s_j is s_i and the corresponding particular r_j^i is r_i^i . The function u^i describes in particular individual i ’s selection of the aspect of each individual j ’s situation she cares for or deems relevant, in depending on these items

while it does not depend on the others. These aspects may depend on the type of relation r_j^i . For example, one may care for the happiness of one's children, and their consumption may only be a means for it, while one may only care for the income and for the respect of basic rights of more distant members of society, and only for the respect of basic rights of still more distant ones, except if they are in a state of deep suffering. And a main function of the r_j^i is the determination of the relative intensity with which individual i cares for the various s_j (this is in particular manifested by the rates of substitution, in function u^i , of comparable evaluated parameters of the various s_j).

The operation of respectful impartialization consists of replacing all r_j^i in u^i by some notional social relation ρ , thus transforming the function u^i into the impartialized function v^i defined as

$$v^i(s, \rho) = u^i(\{s_j, \rho\})$$

where $s = \{s_j\}$ is the set of individuals' situations, symmetrical in the s_j .

The following discussions will argue that the orderings of s defined by the v^i are generally more similar than those defined by the u^i – usually much more so. And yet the functions v^i will generally still represent different orderings, thus failing to yield full consensus or unanimity. Our main topic will consist of ways of building unanimity from functions v^i . We shall also argue that the case where $\rho = r_i^i$ for all i , that is, each individual notionally “loves others as herself”, is bound to be morally particularly relevant. This, however, poses the problem of the definition of the corresponding functions v^i , which will be discussed. And the theory of consensualisation will be worked out in full for this case.

2.2.4 A remark about impartial spectators

Most people can have impartial judgments, in addition to their self-interested and self-centered views and apart from them – at least with respect to others in a certain relation or distance to them. Such judgments can be taken as the individuals' impartial views considered here, though foregoing remarks qualified this possibility. These judgments have often been pointed out. For instance, they constitute one aspect of Adam Smith's “impartial spectator” – one of his central ideas. Both John Stuart Mill and Léon Walras note that people can have such views when they are in a “calm and sedate mood” (though this sometimes happens, on the contrary, in case of particular excitement – generally collective one). Among modern philosophers, Thomas Nagel has particularly emphasized this dual point of view in each individual, who is capable of taking the “view from nowhere”.

However, as we noticed, impartiality is only one property of the structure of an evaluation and it leaves many possibilities open. If individuals can be impartial spectators, there is no guarantee that this would be the same impartial evaluation for each. Hence, the problem of the multiplicity the individuals' impartial views is a priori present with this concept too, unless there is a strong and pervasive social norm defining this impartial judgment. Moreover, there may be a structural relation between an individuals' impartial spectator and her self-centered view. For instance, this impartial or truly impartial view may be one of the notional $v^i(s, \rho)$ defined in the previous paragraph, for some reference relation ρ . This would provide a relation of consistency between the individual's self-centered and impartial views. In addition, some people may lack an impartial or truly impartial view, and

hence resorting to the above notional impartialization is the solution that uses endogeneity. And for individuals who can have such a view, there remains to find out what it precisely is.

This impartial spectator within each individual is a priori something very different from the impartial external observer very often assumed or considered. Such an observer is an arbitrary *deus ex machina* if it is notional, with an a priori arbitrary evaluation. If she is an actual individual, this would violate the moral of endogeneity (and she may not care, or care enough). By contrast, the impartial spectator is a psychological fact of each society's member. However, a notional outside observer or judge may be built from conditions of rationality concerning her evaluation. If society's members are convinced by these conditions, then this evaluation may become each individual's moral or social judgment⁶.

The most intensive use of a concept of an "impartial spectator" is Adam Smith's in his *Theory of Moral Sentiments* (he may have coined the term). This spectator is, for each individual, "the man within the breast", "the great inmate of the breast", "the great demi-god within the breast", "the great judge and arbiter of our conduct". This spectator is the individuals' conscience. Its judgment far exceeds issues of justice, but it includes them since Smith uses the concept of justice in the sense of the Antiquity, as the right attitude toward others. However, there is no emphasis on the aspect of equality or symmetry of impartiality, as is the case with our present concern (universal rights became emphasized in between Smith's time and ours). Nor are we said whether the individuals' impartial spectators tells them the same thing or not – the second of our present concerns. The spectator's impartiality essentially means that it judges with another concern than the interest of the individual it inhabits, and possibly as a critic of her drives. Its scope of concern and its role make Smith's impartial spectator rather akin to Freud's *superego*, but it much differs from it in that it has no mentioned social specificity and origin. There is no reason to doubt Smith's words about his impression of such a host in his chest. But do all chests have one? Do all individuals' impartial spectators judge alike? Or is there, as regards justice, one for Tories who favors natural rights and a more egalitarian one for Whigs? Is there one per type of education, culture, or family? Don't some enjoin one to defend family first, or country, while others take a universal scope and judge all mankind alike? How impartial are the former? Shouldn't we, therefore, rather than take as given the norms of a moral conscience put into us by specific influence, education and formation, build up this judgment with reason as Kant was soon going to forcefully advocate? Smith's notion of the impartial spectator has enabled sociologist Raymond Boudon (1999) to shed light on important social phenomena. This notion, however, is a priori unsufficiently precise for a number of problems for which we would like to use it, such as deciding between equality in incomes (or Rawlsian "primary goods") and the natural right to all the product of one's own capacities, or for finding measures of unjust inequalities. However, economics has largely consisted in making precise Smith's intuitions, over a couple of centuries – such as the efficiency of egoism in exchange (to Pareto and Arrow-Debreu), that of the division of labor for growth, or the problems of monopolies. It may now be time that Smith's impartial spectator be developed for taking normative economics out of its impasses.

2.3 The multiplicity of impartialized evaluations, and solutions.

⁶ For example, Arrow's conditions for a social preference ordering could be of this type. Yet, in this specific case, among various problems, the basic structure which may lead to a solution for facing the problem of distribution and of opposition of self-interest, namely impartiality with respect to self-interests, is absent because the variables necessary for expressing it are not even explicit.

2.3.1 *The remaining variety of individuals' impartialized evaluations*

The (notional) impartialization of the evaluations of society's members is bound to constitute a large progress toward consensus and unanimity. Indeed, it erases the cause of direct, frontal and inherent opposition in individuals' views as concerns the allocation of scarce resources, self-interest and self-centeredness. It leads each individual to consider, equally or on the same footing, the interests, situations, preferences or views of all society's members. Moreover, the possible causes of differences among these impartialized evaluations were also present in the actual views. These causes, nevertheless, a priori exist, and they a priori prevent the considered impartialization from providing full unanimity.

Two issues are present. One is the choice of the social relation of reference retained for the impartialization. People may indeed evaluate and share differently when they consider complete foreigners, common members of the considered society, neighbours, cousins, brothers, children, and so on. However, the same relation of reference should be used for the impartialization of all individuals' views, since there is no reason to discriminate in this respect, and hence for a reason of consistency (and impartiality in the moral use of people's impartialized view). This prevents one possible cause of interpersonal differences in the retained impartialized evaluations, and the choice of this relation of reference will shortly be discussed. There will nevertheless a priori remain another possible reason for differences in individuals' impartialized evaluations: the individuals' specific, *sui generis* preferences or values for evaluation or sharing among people in the chosen social relation of reference with themselves, whatever this relation. Various individuals may evaluate and share differently among individuals belonging to any of the considered categories (brothers, neighbours, common members, and so on).

2.3.2 *The solutions to the problem of the multiplicity of individual impartialized evaluations*

These differences, however, present a number of aspects which are favorable to the derivation of a consensus solution. They may be limited, notably because of the effects of shared social norms. Being, in any event, much more limited than the differences in the actual self-centered individual evaluations, they may permit an easy implementation of other solutions for the social choice (such as an aggregation or a social contract). They are amenable to the same treatment as that applied to the differences in the actual individual evaluations, that is, to impartializations of higher degrees – as we shall shortly see – and consistency may require this solution. Finally, the impartialized evaluations, being moral views (rather than self-interests), are amenable to specific treatments leading to their uniformization, such as moral reasoning and argumentation, notionally erasing the effects of different information in their formation, considering the effects of “empathetic information” (that is, knowing how it feels to have the other's view and not only knowing that she holds it), and “discourse ethic”.

First of all, indeed, the way in which the members of a society judge and share among people who are in a given type of relation with them (close or distant family members, neighbours, other society's members, and so on) is more or less influenced by social norms, habits or customs which are by definition common to a group. This tends to homogenize these views. Note that this does not imply that justice endorses the whole of actual social habits since impartialization can for instance lead one to apply, to all the sharings and transfers in the society, ways of sharing and mutual aid which actually exist among relatives, siblings or neighbours. This role of social norms, however, is absent in the special case we shall particularly focus on, that where the impartialization consists of considering all society's

members as oneself (this case will of course require, to begin with, conceptually defining its possible actual meanings). Moreover, the individuals' following of social norms can be variable and is sometimes absent. In addition, the more heterogeneous the considered society, and hence the larger it is in a sense that implies this differentiation, the less the presence of shared social norms which can provide the solution is likely. Furthermore, if such norms are enforced by social pressure, or otherwise violate individuals' autonomous choice, they might not constitute a legitimate reference. Hence, we have to consider the case where there are different individual impartialized evaluations.

However, we shall see examples of how the impartialization, because it makes individuals' views much more similar than they actually are, provides an easy solution when other methods – notably aggregation – are applied to the impartialized evaluations rather than to the actual ones (section 6).

The most consistent solution, however, consists of noticing that the obtained problem – a multiplicity of individuals' evaluations which may prevent consensus – is formally the same as that which led to the impartialization of actual views in the first place, though progress has been made. Now, consistency suggests or demands that the same problem be faced with the same method of solution. Then, the various impartialized views are considered as the views of individuals that each individual notionally considers as being in the chosen reference type of social relation with her⁷. Each individual arbitrates among these views as she would do among the views of people with whom she is in the considered relation. The new individuals' views are bound to be more similar than the only impartialized individual views they consider because each is formally a kind or average of the latter, or of compromise among them. And yet, there will a priori still be several such views. Then, again, consistency demands that the same problem be faced with the same way of solution. And so on it goes. The process generally converges toward a unanimity of evaluation, which then constitutes the solution. Examples of this moral regress will be provided (section 5)⁸. This method is not actually more complex than one of its steps such as the first one, since it consists of a repetition of it (demanded by consistency).

Finally, apart from the particular case of “treating others as oneself”, which we shall particularly consider, the impartializations consist of expanding the social scope of some actual views, about others, which are not purely self-interested (though they may be self-centered). Then, the differences among these views (or among those of the individuals' own impartial spectators if they exist) can also be notionally erased in erasing differences and insufficiencies in information of various types. Besides present information, this will include the “formative information” received presently or in the past from social environment and life experiences which shape individuals' outlooks and judgments. This will also include individuals' “empathetic information” about others' social or moral sentiments, that is, knowing not only their existence but also “how it feels” to have them with sufficient intensity. (These homogeneization of moral views from information will be closely considered in section 10.3).

2.3.3 *The distance of impartiality*

⁷ A different type of social relation may be chosen for this second level. Note, in particular, that evaluating people's impartialized views and their self-interest (in the first level) are different things. Further levels may raise a similar question. This discussion will not be developed here.

⁸ This is also considered in Kolm 1984.

One parameter of the problem and of the solution is the chosen relation of reference retained for the impartialization of individuals' evaluations. A main aspect of this relation is what sociologists call the "social distance" between individuals. This is here the notional distance between each impartial evaluator and society's members whose situation she evaluates (including herself in this function). This relation and distance have several implications. An individual may have different rules for sharing among individuals who are, for her, foreigners, ordinary community members, neighbours, brothers, children, and so on. This parameter goes along with the intensity of the concern, *ceteris paribus*, and an individual may take greater care of the specific needs or rights of the considered individuals, the more intense her concern about them and hence the closer the relation. It may be, as a result, that the closer the relation, the more comprehensive the set of evaluated items deemed relevant, such as, for increasing distances, happiness (and all its causes), or only consumption or income, or in the end only constitutional rights⁹ (the relation with the structure of the evaluation such as inequality aversion is not so clear). The relation of the (impartialized) evaluator with the evaluated, and the corresponding social distance, also defines the relation and distance of the (notionally) evaluated persons among themselves – they are community members, neighbours, brothers, etc. – in the eyes of the evaluators, and this may influence rules of sharing, possibly as stated by norms of the society. Moreover, the possibility or difficulty in conceiving the impartialization and evaluating what it is, may depend on the chosen relation or distance. Finally, the implementation of rules of justice may be socially and politically more difficult when it requires transfers, solidarities, or rules of sharing which correspond to social relations or distances too different from the actual ones for a larger number of people.

The intensity of concern should at least be positive or exist, since, if not, an opinion provides nothing of the required evaluation. This absence may be the case in considering complete foreigners, and this may be the situation of the "impartial observer" from outside often considered (and possibly of Thomas Nagel's "view from nowhere"). It is likely that the more intense the concern, hence the closer the social relation of reference, the more moral the solution is. Brotherly care is probably better than sharing among foreigners. Specific needs and characteristics will be more carefully taken care of and compared. Then, justice will consist of applying throughout the considered society the comparisons, sharing rules and transfers that exist among people who are socially close to one another. However, while the later rules may be more or less spontaneous in the considered relatively closely knit group, they may have to be supported by the coercion specific to justice for their extension to society as a whole. This may elicit social and political limits to the implementation of justice, though this coercive extension may also in fact be desired, which can be for several possible reasons: Justice in society, or the situation of beneficiaries of the policy, may be desired public goods,¹⁰ and coercion either prevents free riding or guarantees to each contributor that others contribute their fair share; or each individual's impartial spectator or moral self applauds the coercion of her selfish self (coercive justice may act as Ulysses's bonds protecting from following the siren's call of selfishness). When justice actually coercively extends spontaneous solidarities, it is to be seen as a second best to the actual extension to society of the interpersonal sentiments and motivations which lead to these rules and behaviour¹¹. For instance, justice can be the extension, to the whole of society, of sharing and mutual aid which naturally result from brotherly care, reciprocity among neighbours, or family solidarity.

⁹ However, higher concern may lead to more concern for the respect of privacy and free choice of the evaluated persons, which leads to the converse relation between social distance and the items deemed relevant for the evaluation of justice.

¹⁰ See Kolm 1966.

¹¹ This is the issue of "general reciprocity" (see Kolm, 1984).

Moreover, the choice of the notional uniform relation or distance of reference has an impact on the conceptual and informational possibility or difficulty of the operation of impartialization. The alternatives are, for instance, considering an individual's own situation and those of the people she particularly cares for as those of distant members of the society ("objectivization" and "distanciation"), or considering the later as her neighbour, cousin or brother ("increasing concern"). These operations can be made by the ethical analysis and, perhaps, by the individual herself when she takes an "impartial" viewpoint. One cannot say that one of these two opposite operations is always easier or more difficult than the other, conceptually or in transforming one's feelings. However, the particular limiting case of increasing concern to the point of considering others as oneself, which provides one of the most intense concern and the shortest distance and may be the most moral solution, raises a particular conceptual and logical difficulty shortly to be discussed. Solutions to this difficulty will be proposed, but they turn out not to be fully satisfactory. Hence we shall finally return to the discussion of other ethical notional social relations and distances.

2.4 Full concern: *alter as ego*.

2.4.1 Impartial extended joint self-concern.

Indeed, the problem of the conceptual uniformization of the social relation becomes the most acute when one pushes the proximity of the relation – the shortness of the social distance – to its limit, in having each individual consider each other as she considers herself. The person with whom one is the most closely related is in general oneself. The rule of justice then becomes: "care for others as you care for yourself"; if not "love your neighbour as yourself", at least let things be as if it were the case (while the former cases amount to implying "do not love yourself more than you love your neighbour"). This is impartial and full of (notional) concern. This solution may be the most moral one.

This "universal" (or overall) joint empathy has, compared to the previous solutions, one advantage and three drawbacks which are the cost to be paid for its moral quality. The advantage is that the individual evaluates her own situation as she actually sees it. She no longer has to step out of herself in order to see her situation as that of a neighbour or a common member of the society (the objectivization). This, however, is obtained at the cost of the converse problem with a much larger scope since the individual is to be made consider each other as herself for evaluating the overall situation. This raises the three difficulties of *empathy*, *joint empathy*, and *comparative empathy*. First, the individual has to consider being each of the others (and not only seeing herself as some other person). We shall shortly see the various problems this raises. Second, this is to be done jointly while the individual has only one self. This constitutes a basic conceptual problem raised by the pure logic of this new scheme. The problem of split personality of the foregoing solutions is replaced by a problem of multiple personality, which is conceptually worse. Third, the "empathized" situations of all the individuals have to be compared or aggregated by the evaluating individual. But actual life provides no clue for such comparison and aggregation of the situation of individuals which are other oneselves, since the individual only has one self.

Indeed, an individual can have several neighbours, brothers, colleagues, children or ordinary fellow countrymen, and hence one can relatively easily imagine her to be in such a relation with a larger set of individuals, notably for sharing among them. The case of self is a priori different. I may consider others as my *alter egos*. This will imply a few things. I shall

probably be concerned with their end values, possibly their happiness or their satisfaction, rather than by other means, consumption or situation of them, as final concerns. And I will manifest the corresponding impartiality, and equality when the constraints of the problem have the appropriate structure. But what is the solution when equality is not possible, or is dominated by possibilities which make everyone better off, or can be left in giving much more to many at the sole cost of very little less for one, or even in many more ordinary cases? With the previous formalization, each given function u^i is symmetrical in the s_j with the same r_j^i , and this can provide a basis for building the general symmetrical function $v^i(s, \rho)$ with $\rho = r_j^i$; however, for $\rho = r_i^i$, there is only one such r_i^i and s_i , and hence there is no such basis. Hence, a basic problem of the endogenous solution to social choice is:

Problem P or Endogenous Full Concern: Derive a social ethical principle from the views of the concerned individuals with opposed interests, in making each of them as concerned with each other as she actually is with herself – while she only has one self.

This is the problem of finding an *endogenous, fully concerned, impartial social rule*.

Note that this problem is very different from the other, often considered, of an individual having preferences about being someone rather than someone else. Such considerations will have to intervene. But we need much more: we need the individual to consider being jointly all individuals for arbitrating between the individuals' situations in the social choice.

In brief, Thomas Nagel's characterization of impartiality as "the view from nowhere" implies impartiality but is undefined (where is nowhere?). The "view from elsewhere" of the impartial foreign observer may not care and does not exist for society at large. Impartial full concern is, rather, a "view from everywhere", but this demands passing from looking at oneself (egoism) not only to looking everywhere (impartial concern), but also to seeing from everywhere (impartial full concern).

2.4.2 *The solutions: multiple selves, time and uncertainty.*

Solutions of problem *P* can consist of an individual considering that she is each individual as various drives or tendencies of her personality, or successively in time, or as uncertain possibilities. They are respectively the *multiple self*, *moral time-sharing*, and *uncertainty* solutions, the aggregation into one individual as the *compound*, the *succession*, and the *possibility* of the actual individuals. Interpersonal distribution is then reduced to an individual question of, respectively, consumer's choice, saving and borrowing, and insurance. The uncertainty case relates to one of the most famous modern theories in social ethics, the theory of the original position as elaborated by John Rawls. In this theory, the individuals are assumed to evaluate society "in the original position", before they know the actual world and in particular who they will be. The present theory, however, is not of the type proposed and used by Rawls, but, rather, the type he labels "the thin veil of ignorance", associated with John Harsanyi, where the only uncertainty concerns which of the actual individuals one will be. Then, the two latter solutions are respectively a "view from everywhen" and a "view from before".

2.4.3 *Progress and multiplicity.*

The general problems of the endogenous solutions to social choice, and their solutions, noted earlier, will in particular appear with these very specific methods (and their logic will be fully worked out for the theory of the original position). A solution in the endogenous approach to social choice would be obtained if there were unanimity in the impartialized individual evaluations. All the impartializations of individuals' evaluations considered so far will a priori make a progress in this direction, *prima facie* an important one, because all the interests and self-centerednesses of the actual individuals become integrated on a similar footing in each of the new individual evaluations. These evaluations thus no longer present the direct and full opposition of self-interest and self-centeredness, as is the case with the actual individuals' views in problems of distribution of scarce resources. However, they will generally fail to yield full unanimity, because there still is an evaluating individual who evaluates according to her *sui generis* tastes, views or characteristics. As various individuals a priori variously treat their brothers, or their cousins, or their fellow-citizens, a similar thing will occur with the noted solutions to problem *P*. Indeed, various individuals who consider being all actual individuals as alternative uncertain events, successively in time, or as aspects of their personality, will nevertheless have two possible types of reasons for having different evaluations – as it will clearly appear in a forthcoming discussion. They a priori have different preferences toward the dispersion of the desired items in risk, or as concerns evenness or variation of their situation in time, or for facing various drives or competing desires. Moreover, the individuals may a priori have different preferences about being the various individuals.

2.4.4 Solutions to the problem of multiplicity.

Answers to these questions will be proposed, in working out in detail the case of the original position. One solution is in fact implied by the chosen way of solution itself, possibly required by it for a reason of logic and consistency. Indeed, the fact that society's members actually differently evaluate society directly impairs a unanimity solution. Impartializations constitute ways of facing this problem. However, though they provide progress toward consensus, they leave us with precisely the same type of difficulty: a multiplicity of evaluations, one for each evaluator. Then, consistency suggests, or rather requires, that the same problem be faced the same way: in similarly aggregating the various views of the individual evaluators as each individual's possible, successive or multiple views. We will, of course, find a similar result: progress toward unanimity will be made, and yet there will remain a multiplicity of evaluations, one for each evaluator at this second level. Then, consistency again suggests or requires that the same problem be faced the same way, and so on. Each step provides progress toward unanimity. And the process converges toward a generally unique evaluation which constitutes the solution. This iterative solution, repeating the same syntheses at each step, is not really more complex than each step, say than the first synthesis.

Other solutions to the problem of multiplicity will also be proposed, however. They consist of aggregating the impartial evaluators (in the original position for this theory), or of letting them make an exchange agreement in a genuine social contract¹². This provides much more possibilities than the same operations among actual individuals because the impartial evaluators are much more alike to one another. In particular, when the individuals similarly rank the desirability of being the various individuals, they only differ by their preferences

¹² Rawls considers a social contract in the original position, but he also assumes that the individuals in this position all desire the same thing. In an "exchange agreement" individuals a priori disagree but yield something in exchange for others doing the same.

concerning inequality or equality in various uncertain events or at various dates. Then, the logical nature of the considered functions implies specific and simple solutions for the aggregation (see section 6).

2.4.5 The problem of reduction and its solutions

However, the reduction of choices concerning justice to choices in uncertainty, in time, or among one's drives or desires is itself problematic, for reasons which will be presented in section 8 (in the case of the original position). If or when this reduction is not accepted, an endogenous solution to the problem of social choice has to resort to notional individuals' concerns for others that are less than full. For any such solution, however, as we have seen, an initial progress also leaves, a priori, a problem of multiplicity of evaluations – one per individual. Then, we have noted that the same solutions used in the cases of full concern (seeing others as oneself) are also available: regress, aggregation, and social contracts. But we have also seen that other possibilities are then also present, referring to social norms and, in the end, to various aspects of information (this point will be developed in section 10.3).

3. Theories of the original position

3.1 The rationality of the “thin veil” solution

“Assume you had been as unlucky as her” is a common argument for justifying redistribution. Rawls (1971) pushed this standard “intuition” to its limit in entrusting the choice of the principles of justice to the individuals “in the original position”, “behind a veil of ignorance” where they are assumed to know practically nothing of the forthcoming real world, and in particular what individuals they will be¹³. The individuals are then assumed to choose the classical basic rights and the largest possible lowest endowment in primary good. This maximally cautions later choice, added to the protective aspect of the basic rights, implies that sufficiently drastic possibilities are sufficiently likely¹⁴. Rawls, however, only says there is practically full uncertainty. In any event the large assumed introduced uncertainty is *more than necessary for the reason for the theory*, which is *to arbitrate among the interests of known individuals in a known society*. The basic reason for the theory, more precisely, is that the individuals have opposed interests, and hence their views about society are different and opposed. In putting them “in the original position”, facing the same uncertainty, they may have the same opinion (this point will turn out to be crucial). However, it suffices, for this purpose, that each “original” individual does not know which actual individual she will become, but that this uncertainty be equal and full. Any further uncertainty introduced is not justified by the reason for the theory. It is arbitrary (though maximal uncertainty may be a kind of principle). Hence, an uncertainty limited to embodiment into actual people – which Rawls calls the “thin veil of ignorance” – seems more rational, by a kind of moral “Occam's razor” principle, than Rawls' maximal uncertainty – which he labels the “thick veil of ignorance” – (possibly with the pessimistic bias his conclusion requires). The “thin veil” information is used by Harsanyi (1955, 1976, 1977; his 1953 article only considers, as did

¹³ A central distinction is between not knowing what they will be and which of the specific actual individuals they will be.

¹⁴ People commonly choose small risks of very bad events (in using their cars, for instance). Commentators of Rawls usually say that he assumes very risk-averse individuals in the original position. However, individuals in the original position have to have their actual risk-aversion, even though they consider alternatives where they will have different ones – this point will shortly be worked out.

Vickrey (1945), that an individual has an equal probability of obtaining each of the individual *incomes*, a prospect she evaluates with her own ordinary utility for evaluation in uncertainty). It can be added that Rawls does not mention problem *P*, but his own reason for an original position and use of this theory raise questions (they will shortly be noted).

3.2 The multiplicity and individual specificity of the original evaluations and their causes.

3.1.1 The causes of the multiplicity.

Hence, the individuals in the original position evaluate in the uncertainty that they could be or become any of the actual individuals. However, these evaluations will a priori differ from an individual to the other, for several possible reasons. First, the individuals in the original position cannot but face this uncertainty with their own preferences toward risk. Second, the individuals are a priori bound to have different preferences about being various individuals. Third, the individuals may evaluate differently the chances of being or becoming the various individuals.

3.2.2 Equal probabilities

No differences in the chances estimated by the various individuals in the original position will be considered here, because, by construction of the theory and as a result of its intention, each individual in the original position will consider that she has an equal probability of becoming each actual individual¹⁵. Hence, each individual in the original position faces the same prospect. The only differences will be that they a priori evaluate them differently, because they a priori evaluate differently both the alternatives of being the actual individuals and risk.

These equal probabilities of becoming each actual individual result from the nature and intent of the considered theory. An original position is not a description of an actual situation. It is a conceptual ethical theory in social ethics, basically concerned with justice among the actual individuals. This theory consists of an evaluation by the individuals in the original position. In this evaluation, the interests and desires of each actual individual are taken care of by the fact that an individual in the original position considers that she possibly can be or become this actual individual, and only by this fact. Hence, these facts should enter symmetrically in the uncertain prospect of each individual in the original position. With probabilities, this means equal probabilities of being or becoming each actual individual. The reason for this symmetry can be seen in two equivalent ways. It is, as we have seen, the form of impartiality. But we can also directly refer to the logical reason for impartiality. A reason for the value of the probability of becoming one individual applies to all these envisioned actual individuals because no relevant fact can make a difference, by construction of the problem. Hence these probabilities should rationally be equal¹⁶. Their inequality cannot but be unjustified, arbitrary, or resting on an irrelevant fact¹⁷. These probabilities, resulting from reason and ethics, are then normally assumed to be given as objective probabilities to the individuals in the original position. For instance, these individuals receive the same lottery whose prizes are becoming each actual individual, with equal probabilities.¹⁸

¹⁵ Different assumptions are considered by Mongin (1998, 2001) who studies the case with unequal probabilities with a “thin veil of ignorance”.

¹⁶ For a full discussion of the rationality of equality see Kolm, 1998a, new foreword, section 5.

¹⁷ Which could be choosing probabilities by a lottery.

¹⁸ In neighbouring theories, the uncertain event has to be the assignments of the embodiments in the actual individuals to the individuals in the original position (Kolm 1998b, 2001).

An alternative, however, consists of assuming that the individuals in the original position are in complete uncertainty and ignorance about which of the actual individuals they are or will become. Then, with the definition of probabilities based on the principle of Laplace (Condorcet), this very principle says that these probabilities should be considered equal. The fact that these alternatives, mutually exclusive and exhausting the possibilities, are a priori well defined, is favorable to the application of this theory. The impartiality then consists of the same absolute a priori uncertainty in which each individual in the original position is assumed to be concerning which actual individual she is or will become. The crucial moral assumption, however, has been introduced a priori by the choice of these uncertain alternatives. This approach is not deeper, better, superior or more rational than the one of the previous paragraph¹⁹. The only thing which suggests this full uncertainty is that Rawls insists that there is full uncertainty in the original position. Yet, by the same token he rejects the “thin veil of ignorance” and wants the “thick veil”, where the set of possible alternatives is undefined and much larger (in particular, the individuals in the original position can certainly become individuals quite different from the actual ones). Rawls then naturally argues that the uncertainty he considers is not probabilizable. The very non-definition of the alternatives would suffice for this conclusion. Harsanyi (1976) notes complete ignorance (with a “thin veil”) and adds it is satisfied by equal probabilities.

3.2.3 Preferences about being (“ontological preferences”)

The issue of individuals’ preferences about being various individuals raises a number of questions, including logical difficulties, which, however, will be shunt in the proposed solutions to the problems of the theory of the original position (and, more generally, of endogenous full-concern social choice).

Though preferences about being different persons are occasionally expressed, they raise issues of information and of comparison. They also raise the basic issue of who compares when one compares being different persons. And they raise issues of rationality when, for instance, one person prefers to be another one who has the converse preference. Moreover, having preferences about preferences raises the issue of the successive levels of such meta-preferences and of the limitation of these levels. It also raises a problem of *akrasia*, or weakness of the will, when one does not have one’s preferred preferences.

Imagining being someone else is not an uncommon fantasy, in thought experiments, dreams, daydreams, nightmares, and children’s play. Limits to this consideration are also straightforwardly met, as with the common exchange: –“If I were you, I would do otherwise”; – “but you are not me”. The problems of information about what it is and how it feels to be another person are large, but it may be suitable to assume them away – as much as possible – in favour of a hypothesis of perfect information, for the present purpose of pure theory building in ethics. Comparing these various embodiments raise still more difficulties. Yet, people not unfrequently express preferences about being various persons, though usually in casual and spurious remarks rather than thoughtfully. This also applies to preferences between

¹⁹ In fact, the Laplace principle can be given a rational justification (it is then no longer an axiom) identical to the theory of the rationality of equality noted in the previous paragraph (Kolm 1998a, new foreword, section 5, and 1996a, chap.2). This relation is reminiscent of Leibnitz’s proposal to apply the Laplace principle (before Laplace) by analogy with the judicial choice of equality in case of full ignorance. Note that the theory of the rationality of equality is not the age-old and classical “principle of non-sufficient reason”.

being oneself or someone else. This, however, raises particular problems. People who make such comparisons commonly only compare being in the other's place in some sense rather than fully "being" her, thus not substituting all the characteristics (possibly not their "deeper, inner self" – at least they do not know the other's). Indeed, such comparisons raise not only issues of information, but also questions of rationality, logic, and even ontology. A number of issues will be noted, concerning interlocked preferences and in particular reciprocal preferences, the logical and ontological status of self-evaluation, the absence of meta-preferences of higher order, and weakness of the will (*akrasia*).

Can I prefer being person *A* rather than person *B* if *A* prefers being *B* rather than *A* (and one can also assume that *B* also prefers being *B* rather than *A*)? If these preferences implied actual choice, preferring being *A* would be unstable. And there can be cycles and opposite preferences as we shall see. The only stable choices are those of persons who "envy" no one (in this simplistic sense)²⁰. This kind of issues involve questions of meta-preferences, of the "preferring entity", and of preferring and choosing preferences. Tangible hedonistic or eudemonistic evaluations raise similar issues: When I compare the agreeableness of my being *A* or *B*, should I do it with my tastes, with *A*'s tastes, with *B*'s tastes, or possibly with your tastes?

Consider also the case of "reciprocal preferences" where each of two individuals prefers being the other. For instance, if I prefer being you rather than myself and you prefer being me rather than yourself, this implies that I prefer to have – among other things – a preference opposite to my preference, but which, however, implies that I prefer my preference, which itself implies – as we have seen – that I prefer to have the opposite preference, and so on. And you are trapped in the same tricky rationality. The problem is, indeed, much compounded when one introduces third and other persons and preferences between being other persons, with all possible loops. Though all this is possible, it seems that the safest way to quietly live with such a rationality is not to think about it. This, however, opposes an assumption of perfect information that such theoretical constructs legitimately introduce for solving the spurious problems due to informational differences and difficulties (apart from the specifically introduced veil of ignorance).

However, an individual who compares among being various individuals has to make this comparison with some mental apparatus of hers, which she keeps, and hence for which she does not really consider "being in the other's shoes". This seems to be particularly the case when she compares being herself or someone else and expresses preferences about this alternative. If "being someone else" is understood seriously and thoroughly, it may be that no comparison can be made because there remains nothing independent or objective for making the comparison. Scales cannot weigh themselves. There thus is a dilemma. Either the comparison is fully about "being someone" and there is no one for comparing, or there remains some comparing entity and the "being someone" is not complete. For the purpose of ethical theory building, we have to meet this dilemma in accepting the common (imperfect and only partial) view of "being someone else", and hence we shall have preferences about this fact which depend on the comparing individual. These preferences will also depend on how much and which part of herself each individual will keep in comparing, and hence on how much and which part of the people she compares. The only case when this plurality is avoided occurs when there is a sufficient core evaluating system common to all individuals, with which they compare the rest of each individual (which includes the parts of these

²⁰ See Kolm 1971, 1990, 1993, 1995, 2000.

individuals which differ from one to the other). This can happen in a number of cases and situations which will be discussed (notably with a common concept of happiness, success, or the good life, in the considered society – see section 4.2).

Moreover, preferences may imply choice and manifest personalities, and these two types of reasons can lead people to have preferences about preferences (metapreferences). But preferences about metapreferences seem restricted, and preferences of still higher order just seem absent, as we shall see in section 4.1. This, along with other similar issues, limits the possibility of interpersonal comparisons.

In addition, dissatisfaction with one's preferences, or with one's acts that one may consider resulting from preferences ("revealing" them) is weakness of the will or *akrasia*, a kind of lack of rationality (not an uncommon one, though).

However, these logical difficulties, and notably those of the family of interlocked and reciprocal preferences, will be shunt in the proposed solutions to the problem of the original position (and, more generally, of the full-concern endogenous social choice). Indeed, the only aspect of others' preferences which will be used when one considers "being someone else" is her preferences about standard "states of the world" (and not about being someone else). This will suffice because the actual problem concerns choices in this domain only (the selves are actually allocated).

However, we shall also find an interest in the cases where the noted quasi-contradictions of interlocked or reciprocal preferences cannot occur because all individuals order similarly the "being someone" in each given state of the world. This is for instance the case when individuals care only, in the end, for their happiness, their satisfaction, or the goodness of their lives, and they agree on the rankings of these characteristics for all the considered individuals in states of the world. This ordering, often representable by an ordinal function called the "fundamental utility", will not be the case in all possible situations, but it may be in some, and the considered principle of justice is a priori considered for many possible problems (and not solely for the grand problem of a unique universal and intertemporal justice). In these cases, the consideration of the fundamental utility implies the possibility of other solutions, and a particularly clear and meaningful form for all solutions²¹.

3.2.4 *Preferences concerning risk.*

In any event, the individuals' evaluations in the original position will differ because individuals have different preferences concerning risk. This difference is intrinsic to the theory of the original position which consists of introducing a specific uncertainty. The individuals cannot but face this uncertainty of alternatively being several different individuals with their own preferences toward risk. And this aspect of their preferences a priori differs from one individual to the other. Hence, would it only be for this reason, the individuals will have generally different evaluations in the original position. This will be the only reason when there is a fundamental utility. In any event, the individuals cannot help bringing their preferences concerning risk along with them in the original position, if there is an original position.

²¹ See section 4.2 and, for a more complete analysis, Kolm 1998a, new foreword.

Of course, when an individual in the original position envisions the possibility of becoming a particular actual individual, she envisions having in particular her mind and preferences, and notably her *sui generis* preferences concerning risk, which are generally different from those of the considered evaluator. And yet, the evaluation of the risk of having, in real life, these preferences, or others, is made in the original position with the evaluator's own preferences concerning risk. Hence, when an observed actual individual faces risks of real life (not that specific to the original position), the evaluator in the original position first evaluates these risks with the observed individual's preferences about risk. But, this being done, she evaluates the risk of becoming this individual or any other, from the original position, with her own preferences concerning risk.

3.2.5 *The multiplicity*

Hence, there is one evaluation in the original position for each individual. There thus is a multiplicity of evaluations in the original position. Therefore, the theory of the original position, as such, is indeterminate. This opposes both Rawls's and Harsanyi's views for whom all individuals in the original position have the same views. This unanimity is in fact what is sought for in considering a theory of the original position. The aim of this theory, indeed, is to obtain the same view from all different individuals with opposed interests, in placing them in an original position where they are equally uncertain about which actual individual they will be. The uniqueness of the view in the original position solves the problem. Unfortunately, the individuals in the original position, facing this risk (and others with a "thick veil of ignorance"), evaluate it with their preferences concerning risk, which are a priori different from one to the other. And they also often have different preferences about being various individuals (notwithstanding the difficulties raised by this concept).

This multiplicity, however, need not occur, in the end, with Rawls' "thick veil" if one assumes that the worst occurrences are so bad, and also so likely, that everyone prefers to choose the noted maximally protective set of principles whatever her specific tastes in all respects and notably toward risk. And this assumption suits Rawls's aim, which only is to show that *there exists a specification of the risks in the original position which leads to the choice of the principles he advocates*. His use of the theory of the original position is not strictly deductive but only one element in the mobilization of one's considered intuitions in a "reflective equilibrium".

The multiplicity of evaluations in the original position (or "original evaluations"), however, remains with a "thin veil of ignorance", and we have argued that this hypothesis about the original risk has a better claim to rationality than the alternatives, given the actual objective of the theory (facing problem *P*). We shall shortly see both the precise logic of this multiplicity and the ways of obtaining a solution from there. However, one can very easily use one's intuition for grasping a flavor of the properties and solutions.

3.3 *The solutions*

3.3.1 *The progress*

First, though there are several evaluations in the original position – one per individual – rather than only one as one might have hoped, a progress has nevertheless been achieved. Indeed, individuals' evaluations in the original position, though they generally differ, are bound to be more alike to one another than the evaluations of the actual individuals are, because each is a

combination of the evaluations of the actual individuals. For instance, if the actual individuals are purely self-interested, each individual in the original position is jointly concerned with the various self-interests of all the actual individuals she could become. In any even, the set of Pareto-efficient states – the “Pareto set” – will shrink: it will lose elements (in general) and gain none (see section 5.3).

3.3.2 The sequential regress.

Moreover, the logic of the situation is an application of a previous general suggestion. A theory of the original position was used for solving the problem that the actual individuals do not have the same evaluation of society. However, the obtained individuals’ evaluations in the original position still differ. This problem is of the same nature as that one was trying to solve in the first place. And, as we just noticed, though the problem was not solved, a progress was made. Logic, therefore, or at least consistency, require that we face the same type of problem with the same means. That is, we should consider an original position of the original position. This is a deeper, second-degree original position where the risky alternatives, for each individual, consist of becoming each of the individuals in the original position.

The same issues, however, will appear, with the same answers and consequences. For instance, equal probabilities will be chosen for the same reason as above. Moreover, the individuals in the second-order original position still are, as those in the simple original position, the actual individuals as regards their preferences with respect to being someone and to risk. The only differences concern the objects of evaluation: in the original position, they are being the actual individuals; and in the second-degree original position, they are being the individuals in the original position. Therefore, the evaluations of the individuals in the second-degree original position will still differ, for the same two reasons. These individuals face the same risks. But they generally have different preferences toward risk. And they can have different preferences about being the various individuals in the (first-degree) original position. And yet, again for a similar reason, another progress will have been made in that the new set of individuals’ evaluations will still be closer to one another, because each is an aggregate of the same individuals’ evaluation in the original position, and the Pareto set will still shrink.

Then, for the same reason, we should resort, for facing this new multiplicity of evaluations, to another deeper, still anterior, third-degree original position. And so on it goes. At each step, there is an improvement in that individuals’ evaluations become more alike in a sense we shall see (section 5.3) and the Pareto set shrinks. This will generally lead, in the end, to an identity of evaluations and to a single best state. This constitutes the rational solution of the theory of the original position. This infinite, successive repetition of the same process of “originalization” is not really more complicated than each of its similar steps, and in particular than the simple original position. (Section 5 will fully analyze this theory).

3.3.3 An original social contract

However, there are other solutions to the problem of the multiplicity of individuals’ evaluations in the original position. In fact, another solution to the overall problem consists of the individuals making a collective agreement among themselves. However, one may also consider that the individuals in the original position solve the problems raised by their a priori disagreements by making a general collective agreement where each yields something in exchange for the others doing the same. The advantage of agreeing in the original position is

that the evaluations used for making this agreement are more alike than the actual individuals' views are, and hence the domain of a priori disagreement (the Pareto set) is narrower. This hypothetical general agreement would be a kind of theory of the social contract having the original position as "state of nature", the "original social contract" (Rawls sees his theory of the original position as a kind of social contract, but the individuals in his original position a priori all prefer the same choice).

3.3.4 Original aggregations

Still another solution consists of aggregating individuals' evaluations. This is fruitfully done in the original position where the evaluations of the individuals are more similar to one another than their actual evaluations (each is already an aggregation of latter). An "aggregate original position" considers, in the original position, an individual who is an aggregate of the individuals in the original position. With an important specification of the problem shortly to be discussed (the case of a "fundamental utility"), the aggregation is reduced to that of individuals' preferences concerning risk. Then, the structural properties of the necessary notions imply specific meaningful structures for this aggregation (section 6).

4. Extended and fundamental preferences

4.1 Preferences about being

The solutions of endogenous fully concerned social choice, and notably the theory of the original position, lead to the consideration of the individuals considering that they are or could be the various individuals, and having preferences comparing these cases. "Being someone else" is intended to mean having all that concerns her instead of the similar items of oneself. This includes all aspects: the external items such as income, consumption, social situation and relations, and so on, of course; the personal ones such as all physical characteristics; and hence all the mental ones, with, in particular, tastes and preferences. As we have seen in section 3, imagining being someone else and having preferences about such alternatives is not uncommon, though usually not through thoughtful and rational comparison, weighting or empathy. Beyond various problems in information and comparability, such comparisons and preferences raise difficulties in the field of rationality and logic. The comparer can hardly compare her comparing criterion: can one weigh the scales and, if yes, with which scales? The problem of whether, when I compare being yourself and being myself, I should do it with the tastes when I am myself, or when I am you, may not arise because, to begin with, if the comparison is exhaustive it includes these tastes which are no longer available for comparing. Moreover, reciprocal and other interlocked preferences flirt with irrationality, as does the *akrasia* of preferring other preferences to one's own. And metapreferences raise other issues.

Indeed, "being" implies having preferences, and hence having preferences about being implies having preferences about preferences. And if people have preferences about being, they have these second-order preferences. Then, however, being also implies having these latter preferences. Therefore, people also have third-order preferences about them. And so on. The logical problem of self-reference (preferring that I am preferring that I am..., etc) is avoided by the consideration of different levels of preferences. And endless regress does not actually arise because of the psychological fact that the number of these successive levels is limited to two or three, as with all psychological iterations of this type (such as knowing that I know, or that you know with reciprocity, or norms about norms, and so on). As a result, however, the possibility of having preferences about preferences is limited, and so is,

therefore, the possibility of interpersonal comparisons, would it only be for this reason. For instance, I may smoke and prefer not to smoke. But my smoking can be said to “reveal” that I prefer smoking. Hence, I probably prefer not to prefer smoking in this latter sense since these latter preferences lead me to smoke. Can I also prefer not to prefer that I prefer not smoking, possibly for protecting my primary self from intruding (self-)paternalism? Assume also that I prefer rock music to classical music. I may prefer that I prefer classical music, because having such a taste and preference is more classy. And I may also prefer not to have these later metapreference because they are snob. Possibly finding this not to be a good argument does not constitute another level. Further levels seem difficult to grasp. This shows that meta-metapreferences probably constitute a limit, and higher levels of preferences certainly are absent (and they should be considered meaningless rather than resulting from insufficient introspective – or other – information). Hence, there is a level of preferences about which one does not have preferences (and this level is rather close by). This limits the possibility of preferring among various preference systems, and hence about “beings” which imply them.

Hence, “extended” preferences including “being” raise a number of issues in the field of rationality and logic: keeping an evaluating element, reciprocal preferences and related structures, *akrasia*, metapreferences, and others notably as concerns individuals’ information.

Two alternative things will be done with this problem: dodge it and use it. Dodging will result from using, as sole implication of “being someone else”, having this other person’s preferences about the state of the world when she is herself, that is, not using the other person’s preferences about being someone else (or when she is someone else). This will suffice because the actual social choice is only about this state of the world with each person actually being herself. Substitutions of persons are purely notional, and the same result will be obtained whatever is retained in this respect. This will in particular suffice for the increasing similarity of individuals’ preference orderings of the states of the world in the original position, and in the successive original positions in the regress solution. It also permits the two other solutions, original aggregations and an original social contract.

4.2 The case of fundamental preferences or utility.

However, we shall also consider the particular case where the considered individuals in the considered situations similarly order the desirability of being these various individuals²². Such an ordering is called a “fundamental preference”, and, if it is representable by a utility function, this is a “fundamental utility” – an ordinal function (co-ordinal). This structure does not exist in all cases, but it can exist, exactly or with a sufficient approximation, in a number of cases – and the normative theory is for applications for many possible problems (not only once for the the whole world and all times). In fact, problems of justice – our concern here – often arise because some people are clearly less happy than others (and everyone agrees about this fact). This kind of ranking can also have a larger field of application if what is classified is taken to be an objective normative concept in the Aristotalian family of the good life, happiness, flourishing or *eudaemonia*, that the individuals should pursue and may be persuaded to pursue – yet, we shall not pursue this line here. The case with fundamental preferences can have all the properties noted in the preceding paragraph. But it also has a number of further advantages. It provides the specific solutions in the theory of the aggregate original position. It simplifies the other solutions. It is adequate for comparison with classical

²² Individuals’ preferences abide here by this desirability. The term “desirability” suits well the dual aspect of the question: it refers to a subjective observer’s sentiment, and yet the observers see desirability as an objective property of the object (at least all share the “desire” and the corresponding preference).

solutions such as utilitarianism (which has to assume a structure with some similarity and more restrictive because of its cardinality). Finally, the fundamental utility relates to one of the classical pillars of moral philosophy: happiness, *eudaimonia*, or the good life.

If I say that person x is happier than person y , I do not just mean that I would prefer being person x rather than person y , even if I am a eudemonist seeking happiness and with a sufficient inclusiveness of the concept of happiness. I also express something which I feel is objective, and I think I should be able to convince you of this comparison. The notion of happiness has an objective content, even if it refers to a purely psychological fact (as modern uses of the term tend to do). This takes account of the facts that different things make different persons happy (they do not have the same conception of happiness), and that there are various kinds of happiness (such as with serenity or excitement). Moreover, the mere comparison “more” (here “happier”) implies antisymmetry (y is not happier than x), and most of the time also implies transitivity (“ x happier than y ” and “ y happier than z ” imply “ x happier than z ”). Then, happiness has the structure of an ordering. This constitutes the fundamental preferences, a comparison of happinesses, notably across individuals. Then, if all the persons considered are eudemonistic, they prefer being happier, with this understanding of happiness; and their extended preferences, including preferences about being one person or the other, are the fundamental ones, and hence are identical. Eudemonism is a classical and common assumption about individuals (historically favoured by the grand tradition of economics). The foregoing implies only an ordinal concept of preference (hence a “co-ordinal” one). Instead of “happier”, one could have said “has a better life” (which seems still more objective), or, at least in some cases, “more satisfied”.

Note that such interpersonal comparisons, considered as objective, are commonly made. A major obstacle to making them is a lack of information about the individuals, rather than an absence of meaningfulness in the comparison or of agreement in making it. For instance, one cannot say who is happier between two individuals one does know about, and it often happens that one cannot say with a little more information either, while the comparison is clear for a common friend and a common mother. Tinbergen (1957) suggested asking specialists about suffering, such as nurses or medical doctors, which compensation should be given to someone to compensate some cause of pain by comparison with someone else (he does not compare individuals’ happinesses, but this is implied). Courts estimating the *praetium doloris* (the price of pain) for compensations often make interpersonal comparisons of suffering. Scholars who argue that comparisons of happiness can never be made because this is too intimate and personal a concept probably pass by beggars because one cannot say that they are not happier. Now, the difficulties raised by information are irrelevant for our present purpose, since notional full information can – or should – be assumed for the present objective of building a general ethical theory. At least, there is much common use of the terms happy and unhappy, and a happy individual probably is happier than an unhappy one. Let us also add that scholars very often use or imply fundamental preferences or utility as going without saying, for instance when they use a social welfare function, function of individual’s utility levels, which is *symmetrical* in these levels (invariant to permutations) for expressing impartiality, or a maximin or leximin in individuals’ utility²³.

Moreover, the various concepts of happiness, the “good life”, satisfaction, success, flourishing, and so on, often have a cultural dimension and cause, shared by members of a given society, and also influencing these persons’ sentiments, emotions and preferences. This

²³ The leximin is “practical justice” in Kolm 1971.

reinforces the chances of having a fundamental preference, exactly or as a sufficient approximation.

Finally, fundamental preferences preclude the various types of inconsistencies, indeterminacies and irrationalities described earlier (choice of an individuals' preferences for comparing, reciprocal preferences, and so on).

4.3 Rational preferences in uncertainty

The individuals considered will be assumed to evaluate risky prospects according to the classical theory of the rational evaluation of risk, by the mathematical expectation of a specification of their von Neumann-Morgenstern (VNM) cardinal utility function. It is well-known that actual evaluations and behavior often fail to follow this rule. However, the classical properties of rationality of this theory seems to make it appropriate for our present purpose which is to build a normative ethical theory rather than to explain actual behavior or to describe actual evaluations (individual's evaluations may have to be cleaned, laundered, of the corresponding irrationalities).

4.4 Appendix to section 4: Mentions 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 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 not, or is, interpersonally comparable. Harsanyi (1955, 1976, 1977) considers a comparable VNM cardinal utility; this is problematic: the relation between a fundamental comparable utility (notably ordinal) and a VNM utility will be worked out and used in section 5.2. Tinbergen's (1957) discussion implies comparability for "equal happiness" and need no more than ordinalism. Arrow's (1963) 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 5.3, though without solution to the problem of multiplicity of individual evaluations. The ordinal comparability of fundamental preferences has then had a number of uses (Hammond (1976), Arrow (1977), Becker and Stigler (), and others)²⁴. 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).

5. The rational original position solution (the regress)

5.1 Being someone implies preferring like her

²⁴ The normative part of this literature frequently attributes the leximin (or maximin) of "practical justice" to Rawls, while Rawls's first tenet is a rejection of a concept of utility (yet, Rawls (1982) later favourably discusses fundamental utility as introduced in Kolm 1971).

Consider n individuals indexed by i, j , etc. Let x denote a state of the world as classically defined, and $\hat{u}_i(x)$ a specification of the classical ordinal utility function of individual i . Let now $\tilde{u}_i(b_j, x)$ denote a specification of individual i 's ordinal utility function about being individual j when the state of the world is x . Note that the description x includes, among its variables, the consumption, income, social situation, and so on, of each individual, and that “being individual j ” implies having these consumption, income, situation, etc. of individual j in addition to having her tastes, physical and mental characteristics, etc.

The function of x $\tilde{u}_i(b_i, x)$ is ordinally identical to $\hat{u}_i(x)$, and one can choose, as specification for $\hat{u}_i(x)$, $\hat{u}_i(x) = \tilde{u}_i(b_i, x)$. If individual i 's consideration of her being individual j implies that her preference ordering of x is that of individual j , then $\tilde{u}_i(b_j, x)$ as a function of x is ordinally equivalent to the function $\hat{u}_j(x)$, which implies that there exists an increasing $\mathbb{R}^1 \rightarrow \mathbb{R}^1$ function h_{ij} such that

$$\tilde{u}_i(b_j, x) = h_{ij}[\hat{u}_j(x)] = h_{ij} \circ \hat{u}_j(x).$$

(with $h_{ii}=1$).

Let $g_i(\tilde{u}_i)$ denote a specification of individual i 's VNM cardinal utility relevant for her evaluation in risk. Function g_i is increasing. Denote $\hat{v}_i(x) = g_i \circ \hat{u}_i(x)$ ²⁵. In the original position, individual i faces the risky prospect of becoming each individual with an equal probability $1/n$ (see the discussion in section 3.2.2). Her evaluation of this prospect is her expected utility

$$\begin{aligned} \hat{v}_i^1(x) &= n^{-1} \sum_j g_i \circ \tilde{u}_i(b_j, x) = n^{-1} \sum_j g_i \circ h_{ij} \circ \hat{u}_j(x) = n^{-1} \sum_j \gamma_{ij} \circ \hat{u}_j(x) \\ &= n^{-1} \sum_j g_i \circ h_{ij} \circ g_j^{-1} \circ \hat{v}_j(x) = n^{-1} \sum_j H_{ij} \circ \hat{v}_j(x), \end{aligned}$$

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.

These functions represent the n orderings of the x of the individuals in the original position. They are a priori different. However, since the functions γ_{ij} or H_{ij} are increasing, the orderings of the x defined by the \hat{v}_i^1 are generally more alike than those defined by the \hat{v}_i in a sense shortly to be made precise and which in general implies a shrinking of the set of Pareto-efficient states – the “Pareto set” (section 5-3). Hence, the problem of obtaining a unique

²⁵ We can make precise the relation among the various utilities. Denote as $(\beta, \xi) = \{\beta^k, \xi^k; p^k\}$ the risky prospect where (b_j, x) take the values (β^k, ξ^k) with probability p^k ($\sum p^k = 1$ and $p^k \geq 0$ for all k). Denote as $U_i(\beta, \xi)$ a specification of individual i 's ordinal utility function evaluating (β, ξ) , such that $U_i = \tilde{u}_i(b_j, x)$ when (β, ξ) becomes the sure prospect (b_j, x) . With the classical rational evaluation of risk, there are two functions $G_i: \mathbb{R}_1 \rightarrow \mathbb{R}_1$ increasing and $\tilde{v}_i(\beta^k, \xi^k)$ such that $U_i(\beta, \xi) = G_i[\sum_k p^k \tilde{v}_i(\beta^k, \xi^k)]$, where \tilde{v}_i is a specification of the corresponding VNM cardinal utility of individual i . In particular, when (β, ξ) is the sure (b_j, x) , this relation becomes $\tilde{u}_i(b_j, x) = G_i \circ \tilde{v}_i(b_j, x)$, or, denoting $g_i = G_i^{-1}$, $\tilde{v}_i(b_j, x) = g_i \circ \tilde{u}_i(b_j, x)$. When $j=i$, this becomes $g_i \circ \hat{u}_i(x) = \hat{v}_i(x) = \tilde{v}_i(b_i, x)$, and more generally $\tilde{v}_i(b_j, x) = g_i \circ \tilde{u}_i(b_j, x) = g_i \circ h_{ij} \circ \hat{u}_j(x) = g_i \circ h_{ij} \circ g_j^{-1} \circ \hat{v}_j(x)$. Then, the theory of the original position considers the uncertain prospect $(\beta, \xi) = \{b_j, x; 1/n\}$.

ordering unanimously agreed upon is not solved, but a progress has been made in this direction.

Moreover, the problem of having one evaluation per individual in the original position is analogous to the initial problem in the real world. Consistency requires one to face 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 \hat{v}_i^1 with the same probability $1/n$. Then, individual i 's evaluation in this second-degree original position $\hat{v}_i^2(x)$ obtains from the $\hat{v}_j^1(x)$ as the latter were obtained from the $\hat{v}_k(x)$, that is,

$$\hat{v}_i^2(x) = n^{-1} \sum_j H_{ij} \circ \hat{v}_j^1(x),$$

There still are n evaluations. However, for the same reason as above, the orderings of the x defined by the $\hat{v}_i^2(x)$ with generally be more alike than those defined by the $\hat{v}_i^1(x)$. The process than has to be repeated, in successively anterior original positions $OP_1, OP_2, \dots, OP_m, \dots$, with the recurrence

$$\hat{v}_i^{m+1}(x) = n^{-1} \sum_j H_{ij} \circ \hat{v}_j^m(x),$$

and the orderings defined by the $\hat{v}_i^{m+1}(x)$ being generally more alike than those defined by the $\hat{v}_j^m(x)$, with in general a shrinking of the Pareto set.

5.2 The case of a fundamental utility

When there is, or there can be assumed to be with a sufficient approximation, a fundamental utility for the considered individuals and problem (situations), the two aspects of this concept can be considered: a common ordering of the desirability of being such individual in such state of the world, and an “objective” conception of happiness, satisfaction, success, or the goodness of life. With variables defined as in the foregoing section, the individuals have a common preference ordering, or ranking in terms of desirability which they share, of the items “being individual i in state of the world x ” ($b_{i,x}$). This is 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 , $u_i(x) = u(b_{i,x})$ is a specification of individual i 's ordinal utility function for evaluating x . The other specifications of the ordinal fundamental utility are $\varphi \circ u(b_{i,x}) = \varphi \circ u_i(x)$, where φ is any increasing function. Hence, the functions $u_i(x)$ can be replaced by any functions $\varphi \circ u_i(x)$ with the same function φ for all i : they are co-ordinal. But the other specifications of individual i 's ordinal utility are $\varphi_i \circ u_i(x)$ where φ_i is any increasing function (which can depend on individual i).

A specification of individual i 's cardinal VNM utility is

$$f_i \circ u_i = f_i \circ u(b_{i,x}) = f_i \circ u_i(x) = v_i(x).$$

where f_i is an appropriate increasing function (its cardinality says that it can be replaced by any function $a f_i + b_i$ where a_i and b_i are constant and $a_i > 0$).²⁶

²⁶ We can make precise the relations among the various utilities. A specification of individual i 's ordinal utility for evaluating the uncertain prospect $(\beta, \xi) = \{\beta^k, \xi^k; p^k\}$ (see note 25) is $U_i(\beta, \xi) = F_i[\sum_k p^k v_i(\beta^k, \xi^k)]$ where v_i is a specification of individual i 's VNM cardinal utility and F_i is a $\mathfrak{R}_1 \rightarrow \mathfrak{R}_1$ increasing function. Choose this U_i which is the chosen specification of the fundamental utility when (β, ξ) is certain: when (β, ξ) is the certain $(b_{j,x})$, $U_i(b_{j,x}; 1) = u(b_{j,x})$ (this determines the function F_i).

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:

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

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

$$u_i^1(x) = f_i^{-1} \circ v_i^1(x) = f_i^{-1}[n^{-1} \sum_j f_i \circ u_j(x)],$$

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

$$u_i^1(x) = M[\{u_j(x)\}, f_i].$$

Since this case is a subcase of that of the previous section, 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 u_i^1 are more alike than those implied by the u_i , and the Pareto set shrinks. 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

$$u_i^{m+1}(x) = M[\{u_j^m(x)\}, f_i]$$

in fundamental utility and

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

for the VNM utilities. 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 $m \rightarrow \infty$ and $u_i^m \rightarrow u_i^\infty$, the u_i^∞ satisfy

$$u_i^\infty(x) = M[\{u_j^\infty(x)\}, f_i]$$

or

$$n f_i \circ u_i^\infty(x) = \sum_j f_i \circ u_j^\infty(x),$$

for all i . These conditions are satisfied by any set of equal levels u_i^∞ for all i . For all x , they are satisfied by any equal functions $u_i^\infty(x) = U(x)$. This condition of an equal level of happiness or satisfaction (in fundamental utility) is “eudemonistic justice”²⁷. However, this is for the individuals in the “infinite original position”, not for the actual individuals.

Then, $F_i[v_i(b_j, x)] = u(b_j, x) = u_j(x)$, or, denoting $f_i = F_i^{-1}$, $v_i(b_j, x) = f_i \circ u_j(x)$. One can in particular denote $v_i(x) = v_i(b_i, x) = f_i \circ u_i(x)$. The theory of the original position takes $(\beta, \xi) = \{b_j, x; 1/n\}$. Note the dual roles of functions u_i (fundamental) and v_i (VNM): functions u_i can meaningfully be compared by \geq (and hence functions $\varphi \circ u_i$ with any increasing function φ also can), but functions v_i cannot; on the other hand, values of u_i cannot be meaningfully be added, while values of v_i for the same i can. Of course if another specification of the fundamental utility is chosen, say $\varphi(u)$, functions f_i incur the corresponding contravariant transformation in becoming $f_i \circ \varphi^{-1}$.

²⁷ See Kolm 1971.

From its construction, $U(x)=W[\{u_i(x)\}]$. Function W is an increasing symmetrical function of the u_i (at each step, each u_i^{m+1} is an increasing symmetrical function of the u_i^m). This increasingness guarantees Pareto efficiency of the result. If all individuals are very risk-averse, $u_i^1(x) = u_i^\infty(x) = U(x) = \text{Min}_i u_i(x)$, which is eudemonistic “practical justice”²⁸. Note that the direct equality of the $u_i(x)$ may have to violate Pareto efficiency, or may not be possible. This was one reason for resorting to practical justice. But this solution was too extreme for a general solution²⁹. If all functions f_i were the same (cardinally, that is, up to an affine increasing function) and were function f , then, for all i and $m \geq 1$, $U(x) = u_i^m(x) = u_i^1(x) = M[\{u_i(x)\}, f]$ and a maximand can be $nf \circ U(x) = \sum f \circ u_i(x) = \sum v_i(x)$ in calling $f(u_i) = v_i$, a form usually associated with Harsanyi. However, there is no a priori reason for the identity of the functions f_i or v_i . Functions f_i depend on the characteristics of individual i , but this is implied by the relation between the fundamental comparable u_i and the specific VNM v_i (see note 26).

³⁰

5.3 Homogeneization and convergence of individual preference orderings

The way to general consensus and unanimity in the sequence of original positions has, more precisely, the following structure. Denote as $s_i(x)$ and $t_i(x)$ two sets of n functions of x , each function representing an ordering of the x , and consider n increasing functions S_i defining the s_i from the t_i : $s_i(x) = S_i[\{t_j(x)\}]$. The concept to be shown is that the orderings represented by the s_i become more alike, or at least no less alike, than those represented by the t_i . Denote the vectors $s = \{s_i\}$ and $t = \{t_i\}$. Use the usual vector inequalities: for vectors $\tau = \{\tau_i\}$ and $\tau' = \{\tau'_i\}$, $\tau \geq \tau'$ is $\tau \geq \tau'_i$ for all i , $\tau > \tau'$ is $\tau > \tau'_i$ for all i , and $\tau \geq \tau'$ is $\tau \geq \tau'_i$ for all i and $\tau > \tau'_i$ for at least one i . Then, for x and x' , $t(x) \geq t(x') \Rightarrow s(x) > s(x')$. Hence, $t(x) \geq t(x') \Rightarrow s(x) \geq s(x')$. Also, $t(x) \geq t(x') \Rightarrow s(x) \geq s(x')$. This means that weak unanimity (\geq), strong unanimity ($>$), and “Pareto unanimity” (\geq) are maintained in the transformation. That is, the sets of x' which are unanimously preferred to a given x , or to which a given x is unanimously preferred, according to any of these three unanimities, lose no element in the transformation. Generally, they will expand.

Moreover, given a set of orderings represented by functions $\tau_i(x)$, denote as $P(\tau)$ the “Pareto set” of the x – the set of Pareto-efficient x –: if Π is the set of possible x , $P(\tau) \subseteq \Pi$, and $x \in P(\tau) \Leftrightarrow x \in \Pi$ and $\tau(x') \geq \tau(x)$ for no $x' \in \Pi$. Then, $P(s) \subseteq P(t)$. Indeed, $x \in \Pi - P(t)$ implies that there exists a $x' \in \Pi$ such that $t(x') \geq t(x)$. But this implies $s(x') > s(x)$ and hence $s(x') \geq$

²⁸ Id. Practical justice is more generally the leximin in the $u_i(x)$.

²⁹ This remark is the last sentence of this work (Kolm 1971).

³⁰ Mongin (2001) provides an extensive exegesis of Harsanyi’s thought about this question (this is not our present topic). Pattanaik’s (1971) discussion amounts to saying that there necessarily is a fundamental utility in certainty and in taking x as certain, but that the individuals’ VNM utilities can differ. The present regress solution can be seen as rescuing Harsanyi’s intent from a basic problem, or as implementing what he was looking for by an iterative regressive application of his original intuition.

$s(x)$. Therefore, $x \notin P(s)$. And hence $P(s) \subseteq P(t)$. That is, the Pareto set loses no element in the transformation. It generally shrinks³¹.

These results apply to the successive set of functions $\hat{v}^m(x) = \{\hat{v}_i^m(x)\}$, $u^m(x) = \{u_i^m(x)\}$, and $v^m(x) = \{v_i^m(x)\}$ of the two previous sections, in noting respectively $\hat{v}_i(x) = \hat{v}_i^o(x)$, $u_i(x) = u_i^o(x)$, and $v_i(x) = v_i^o(x)$, for passing from the set m to the set $m+1$ for all integers m from 0 on. In particular, denoting as P^m the Pareto set for the set of functions of order m in each series, $P^o \supseteq P^1 \supseteq \dots \supseteq P^m \supseteq P^{m+1} \supseteq \dots$.

6. The aggregate fundamental original position.

Another, natural, way of facing the multiplicity of individual evaluations in the original position consists of aggregating them. One could also, of course, directly aggregate individuals' actual evaluations. However, the theory of the original position provides the possibility of particular remarkable specific forms of aggregation and reasons for them. The theory of the aggregate original position chooses, as evaluation, that of an individual in the original position who is a notional aggregate of the actual individuals (in the original position). The specifications are particularly remarkable when a fundamental utility exists or can be assumed. Then, the only characteristics to aggregate are individuals' preferences concerning risk. They are $\mathfrak{R}_1 \rightarrow \mathfrak{R}_1$ functions. Moreover, we may want an aggregation immune from the arbitrary characteristics of the used functions (cardinal for the VNM utility functions and ordinal for the corresponding common fundamental utility). Finally, a principle of independence of irrelevant alternatives will lead to a simple and meaningful family of solutions.

The aggregate evaluator in the original position considers she can have each of the individual's values of a chosen specification of the fundamental utility, $u_i(x)$ for individual i , when the state of the world is x , with equal probabilities. She is an expected utility maximizer with the VNM utility $f(u): \mathfrak{R}_1 \rightarrow \mathfrak{R}_1$ for this risk. Her VNM expected utility is

$$v(x) = n^{-1} \sum f \circ u_i(x) \quad ,$$

and her evaluation of x is $v(x)$ or $V(x) = n v(x)$.

The corresponding aggregation then consists of aggregating the functions f_i into the function f . Hence, function $f(u)$ is a priori a functional of the n functions $f_i(u)$. However, functions f_i and f are cardinal – defined each up to its own arbitrary affine increasing transformation –, and function u , here the variable, is ordinal – defined up an arbitrary increasing transformation, the same for all functions f_i . If we find a justified specific form for the relation giving a specification of f from specifications of f_i , it may not be maintained under these transformations of the f_i , f , and u . Hence, justified forms for this relation are among those which are maintained by these transformations.

First, a cardinal function $\varphi(u)$, which thus can be replaced by any $\alpha\varphi(u) + \beta$ with constant α and β (and $\alpha > 0$) is characterized, if it is twice differentiable, by the function $R(u) = \varphi'' / \varphi'$ which does not depend on this transformation and integrates into $\alpha\varphi + \beta$ with arbitrary

³¹ Properties of this type have been noted by Edgeworth (1888) for two individuals and, a contract curve and additive combinations, and in general by Kolm (1963) and Archibald and Donaldson (1979).

α and β : function $R(u)$ is a sufficient functional invariant of cardinal function ϕ . For functions f_i and f , assumed to be twice differentiable, denote $r_i(u) = -f_i''(u)/f_i'(u)$ and $r(u) = -f''(u)/f'(u)$.³² The aggregation then should be that of functions r_i into function r .

We now add that the functional $\{r_i\} \rightarrow r$ takes the simple form of the same function for each u , $r(u) = \phi[\{r_i(u)\}]$, a “uniform independence of irrelevant alternatives” (an alternative is here a level of u and hence of the $r_i(u)$, and $r(u)$ depends only on these latter levels). Function ϕ will be symmetrical because nothing else relevant differentiates the individuals. It will be increasing for positive sensitivity to individuals’ views, or at least non-decreasing and increasing in one r_i . And if all r_i are equal, their value will also be r , as a respect of individuals’ views implying a respect of unanimity.

The ordinality of u means that it can be replaced by any function $t(u)$ where t is an increasing function. Then, a function $\phi(u)$ becomes $\psi(t)$ with $\psi[t(u)] = \phi(u)$. This implies, for twice differentiable functions, $\phi''/\phi' = t' \cdot (\psi''/\psi') + t''/t'$. The increasingness of $t(u)$, its only required property, permits any positive value for t' and any value for t''/t' . Then, such a change of u simultaneously transforms in this manner functions f_i and f , and hence $r_i(u)$ is transformed into $ar_i + b$, and $r(u)$ is transformed into $ar + b$ with $a = 1/t'$ and $b = t''/t'^2$. This b can have any value, and this a can have any positive value. If function ϕ is to be independent of the chosen specification of ordinal function u , then any transformation of all r_i into $ar_i + b$ has to transform r into $ar + b$, that is, $\phi(\{ar_i + b\}) = ar + b$ for all b and all $a > 0$.

Then, all the noted properties of function ϕ imply that it is of the form $\phi(\{r_i\}) = \sum a_\rho r_\rho$ with constant a_ρ such that $a_\rho \geq 0$ for all ρ , $\sum a_\rho = 1$, and $\{r_\rho\}$ is the ordered permutation of the r_i , that is: the set of indexes ρ is a permutation of the set of indexes i , and $\rho > \rho'$ implies $r_\rho \geq r_{\rho'}$.³³ Particular cases are r which are: the average $r = (1/n)\sum r_i$, the lowest $r = \min r_i$, the median, the average of the r_i in the lowest quantile for some quantile size, or some set of a_ρ satisfying $a_\rho < a_{\rho'}$ when $\rho > \rho'$.

7. A fundamental social contract

Rawls presents his theory of the original position (very different from the one considered here) as a theory of the social contract. In classical theories of the social contract, the individuals in a notional “state of nature”, with their different preferences and characteristics, jointly make a unanimous collective agreement which becomes a joint voluntary contract, and this becomes the moral rule in the real world, which could legitimately be enforced upon the individuals, and with which, indeed, they would have a moral obligation to comply. The most particular feature of Rawls’s original-position social contract is that the individuals in the original position, “behind the veil of ignorance”, all a priori prefer the same rule (which is what Rawls demands from them), in the absence of any agreement. We have seen that this cannot result from their having the same overall preferences in the original position (they have to bring at least their own risk-aversion along with them), but one can imagine a structure of risks such that they will all make the same choice (of Rawls’s principles), and this

³² The signs of the r_i and r depend of the specification chosen for u and hence have a priori no meaning.

³³ This permutation $\{r_i\} \rightarrow \{r_\rho\}$ permits function ϕ to be symmetrical. The other noted property requires that it has the appropriate piecemeal linearity.

suffices for Rawls. Then, these original individuals do not have to and cannot make an exchange-agreement in yielding something from their first best in exchange for others doing the same. They only have to, and can only, take their common desire as the text of the contract and all sign it, in order that it becomes the social moral rule in the real world (Rawls does not discuss that, but one may consider that it is implicit in calling his original position a theory of a social contract). Harsanyi suggests no contract of any kind.

However, in a “thin veil” original position, we have seen that the individuals neither have the same preferences about the alternatives and risk, nor a priori have the same best choice. And yet, one may see the individuals in an original position with different preferences v_i^1 as people in a classical state of nature. Then, they can make an exchange-agreement about future policy, each yielding from her first best in exchange for others doing the same (possibly in choosing a maximand function $V(x)$), and this becomes the social contract to be enforced (and even a moral obligation). Then, any relevant concept of arbitrage or of bargaining provides the solution.³⁴ An advantage of considering a social contract among the individuals in the original position (with preferences v_i^1 or u_i^1) rather than directly (with preferences u_i) is that they already agree much more, the domain of disagreement is much more restricted – this would still more be the case if anterior original positions (as in section 5) were chosen for making the social contract. We shall not pursue this idea further here.

8. Differences between risk and justice³⁵

8.1 *Reducing justice to risk*

However, a very basic question mars or limits the validity of theories of the original position as ethical theories. A theory of the original position reduces a choice in social ethics, which determines distributive justice, to self-interested individual choices in uncertainty. It certainly is very agreeable to find a solution to the difficult problem of social ethics and justice in reducing it to a question of a purely individual choice concerning the chooser’s interest only, about which we have much information from the analysis of individual rationality, psychology, experiments, introspection, and varied literatures. This reduction, however, is questionable and at least has limits. The basic point is that one has the right to take a risk for oneself, while no one has the right to commit an injustice. One is responsible for one’s choice concerning oneself only, while a choice of justice is accountable toward others, society, possibly morals, that is, to something which transcends at least any individual chooser. A choice of justice is a duty, even when it concerns individuals’ final values such as happiness (which is often not its direct or explicit concern), while an individual self-concerned choice in uncertainty only considers consequences (setting aside the particular emotions which the pure fact of uncertainty or risk can raise, which are something else). Risk is chosen by tastes, justice is by norms. The two types of choices have very different natures and rationales. They are different worlds. They have no a priori reason to coincide when their formal setting is made comparable.

One way of reducing justice to individual risk-taking consists of considering a theory of the original position as a theory of the social contract, or as integrated in such a theory. This is what Rawls does. And this can be done with other assumed original ignorance and

³⁴ The possible theories of arbitrage or bargaining are gathered in Kolm 1996a, chap.13.

³⁵ This section develops a critique of the theory of the original position presented in Kolm 1996a, chap.8 (and 2).

whether the individuals in the original position have identical or different preferences or best choices. Then, however, the question is pushed back to the justification of a concept of a social contract and of this particular one.

The differences between choices of justice and individual choices in uncertainty appear about all aspects of the choice when they are sufficiently considered. They appear, first, at each stage of the choice: the rationales, the choice itself, and the result. They also appear about various essential and related aspects such as the nature of the choice, the question of preferences for equality or inequality among uncertain alternatives or among individuals, the chooser's responsibility, or the nature of the relevant items.

8.2 The problems with utilitarianism and welfarism

The foregoing variants of the theory of the original position lead to social ethical evaluation functions depending on individuals' utilities, sometimes in an additive form:

$$W[\{\hat{u}_i(x)\}], \Sigma f[u_i(x)], \Sigma_j H_{ij}[\hat{u}_j(x)], \Sigma_j f_i[u_j(x)], \text{ etc.}$$

They are "welfarist" maximands (as John Hicks says). The additive forms add specifications of individuals' utility functions. They are not utilitarianism in the strict sense since they cannot be said to add individuals' happinesses, but they have one aspect of utilitarianism, and their addition is logically valid (they are "utilitaromorphisms").³⁶ Hence, a number of the criticisms applied to utilitarianism and to welfarism apply to these results, and, consequently, to the theory that derives them. These criticisms are the object of an abundant literature and hence cannot be repeated here. Then, two conclusions can be drawn: either the theory of the original position constitutes an answer to these criticisms, or these criticisms are carried over to this theory and, therefore, to the reduction of the social ethical choice to an individual choice in uncertainty. Conversely, the issues and problems we may find with this reduction or assimilation can be straightforwardly carried over into the debate about utilitarianism.

We shall first propose a few examples in the form of questions, and then try to analytically consider the issue.

8.3 Distributive justice and individual risks: examples of comparisons

The issue of the possibility of reducing justice to risk leads one to ponder the validity of the analogy with questions such as the following ones.

Risk: You have the right to take your car for saving ten minutes walking in going somewhere, at the risk of a fatal traffic accident with probability 10^{-6} . Justice: Is it right to kill someone in order that 999.999 persons save ten minutes? (A worker who may die from accident in building a tunnel which may enable drivers to gain ten minutes will have freely chosen this risk).

Risk: You have the right to bet all your money on a single horse in order to try your luck of knowing how it feels to be a millionaire. Justice: is it right to give all the money to a single person in order that there be a millionaire's greatly pleasurable experience? Similarly, unobjectionable buying of lotteries tickets may amount to morally problematic distributions in the analogy.

³⁶ See Kolm 1996, chap. 14.

Risk: Assume you can have either ten dollars for sure, or a risk of having 0 or x with equal probability; then, you choose the risk if $x > x_r$ for some value x_r . **Justice:** Assume two persons can have either ten dollars each, or one can have x while the other has 0; it is found good to “sacrifice” one in giving her 0 if the other can have at least $x > x_j$ for some value x_j (x_j may be infinite). Then, why should we have $x_j = x_r$? One impression is that we should have $x_j > x_r$ because, in the case of risk, the individual is responsible for choosing her risk of having zero, while, for justice, level 0 is imposed on one person. Putting this otherwise, it seems that, for the same dispersion in interpersonal distribution and in a personal risk, the moral equal equivalent should normally be lower than the certainty equivalent. That is, it seems that moral inequality-aversion should exceed risk aversion. This would be for the judgments in uncertainty or of justice of the same person (the second would be the judgment of the “impartial spectator within her breast”).³⁷ In particular, if all the considered individuals have the same risk-aversion, justice among them would be more inequality-averse.

The difference between risk and justice can affect not only the dispersion but also the nature of the items. For instance, an individual fully evaluating a financial risk will in the end consider what she would do with her various alternative incomes. In a choice of distributive justice, by contrast, each beneficiary may be responsible for how she spends her income, and accountable for the pleasure she derives from it because it is a private matter irrelevant for the public issue of justice (except for exceptional cases of pathological deep depressions).

8.4 Nature of the choices and of the differences

An individual has the right to take a risk for herself, while no one has the right to commit an injustice. A basic consequence of rationality in the basic sense of “for a reason” is a requirement of *prima facie* equality in the relevant items in cases identical in the relevant characteristics.³⁸ The “cases” are either individuals (say) for justice, or alternatives for risk – and then identical characteristics imply equal chances (say, probabilities). For risk, equality means certainty – which a priori has nothing to do with justice as regards their nature. For making the comparison possibly relevant, we have to discard all the particular sentiments and emotions which are the particular and specific companions of uncertainty: the ex ante fear, hope, anxiety, anguish, or excitement due to adrenaline flow, the ex post regret or its opposite, and the religious-like notions which are their frequent associates. There will thus only remain the evaluation of the bundles of alternatives. For justice, by contrast, we certainly cannot eliminate the “sense of justice”, or even the indignation toward injustice, since they seem necessary for comparing the possibilities. Then, in both problems, a reason for deviating from equality is that the gains in some cases may be seen as overcompensating the losses in others (or, possibly, the gains in all cases may compensate deviations from equality or certainty). This, per se, raises no particular issue for the individual choice in uncertainty. This is just as choosing less bread for more butter. There is no reason to stop short of the marginal equalities of the economics textbooks. This will happen with any kind of structure of individuals’ preferences about the set of values. In particular, an individual’s taste for high values may make her favour inequality in this set. This can happen independently of the above noted particular sentiments raised by risk and uncertainty, for actuarially fair risks or for others, with or without convexities in the VNM utility function. In the case of interpersonal distribution, on the other hand, there seems to be a particular responsibility toward the

³⁷ For instance, there can be moral inequality-aversion, and yet the individual may prefer some risk in earnings rather than their actuarial value with certainty (convexity of her VNM utility).

³⁸ See Kolm 1998a, new foreword, section 5.

individuals who lose, which should at least restrain the movement. The result will be in between equality or maximin (full responsibility toward the losers), and the highest sum of the compared items (such as utilitarianism if they intend to be individuals' happiness).

The responsibility of justice toward the losers is bound to be more acute when what is at stake is more important, especially when it is vital. People's increasing cautiousness does not seem to match. We noted that people take risks of death or serious injury which cannot be imposed on particular people as just or acceptable social ethics (though cases of medical care or war raise such issues). Cases of justice about goods of secondary importance may be close to duplicating individual choice in risk. General income distribution should be an intermediate case (if dire misery is taken care of otherwise).

8.5 Different relevant items

The difference between a self-interested choice and a choice of justice also affects the nature of the individuals' items considered relevant, notably for direct consideration in an evaluation function. In the individual choice in uncertainty, the individual is concerned with her own overall final objectives, notably those represented by her utility level. The choice of justice, by contrast, commonly focusses on other items, in neglecting (treating with "benign neglect") a domain which is deemed the private sphere of individuals' choices and evaluations, for reasons of privacy or of freedom of choice and responsibility. For instance, for respectful individualistic justice³⁹ the relevant items can be means of individuals' choices or evaluations, such as income, overall or specific consumption, the satisfaction of certain needs, rights, freedoms, powers, and so on, rather than some "utility". Such views are for instance revealed by complaints about inequalities in these items (certain rights, powers, consumption goods, the satisfaction of particular needs, incomes, and so on). This question of the nature of the relevant item also rejoins – not surprisingly – a usual objection to utilitarianism.

8.6 Different phenomenologies of choice

All the foregoing differences, and all the specific reasons that influence choices of justice and choices in self-interested uncertainty, lead to a different phenomenology of choosing in both cases. This shows in the criteria for choosing and pondering, in their reasons, nature and intensity. There also are similarities. For instance, the property of "separability", in the sense that allocating among two uncertain alternatives or two individuals should not depend on other alternatives or individuals' situation, has good reasons both as concerns risk and justice. Yet, the basic differences in the preferences concerning equality or inequality, or the nature of the equalized items, will intervene. The comparison will not be pursued here, but one can notice that empirical investigations using questionnaires have shown a number of differences among actual choices or preferences in uncertainty and concerning distributive justice (Amiel and Cowell, 1998).

8.7 Responsibility and equality in the original position

8.7.1 Brute luck and option luck

If individuals could freely choose the lotterie ticket corresponding to the original position (thin veil), and then are also given the policy tools which enable them to adjust the situation

³⁹ See Kolm 1996, chap.1.

of the various outcomes – being the actual individuals –, then the respect of their freedom would endorse their lotteries and policy choices and declare the overall situation socially good and just. In fact, the individuals cannot choose the lottery of the original position. If the person they are is seen as “luck”, then it is brute luck and not option luck as when they have the option of taking the risk or not (and have sufficient information about it and make a careful choice). The outcomes can more or less add, to this brute luck, the effects of the policy preferred by the individual, but this only is one of the causes. The individuals cannot be held responsible for the brute luck part of the causes, since, actually, they have not chosen this risk. In particular, they may not have accepted the bad outcomes (brut bad luck), whose relief by the policy may be costly in terms of the other situations and hence may have to be limited. This is the same issue of responsibility as previously noted. This remark also amounts to applying, to the theory of the original position, the objection to the other Harsanyi’s (1955) justification of his utilitarian form.⁴⁰

8.7.2 Equality

The individuals in the original position face identical prospects. This constitutes an equality. Hence, an ethical theory of the original position asserts that this equality is the one proper for justice, the just equality. However, from the original position, the original lottery has to be drawn, for obtaining the actual world. Saying that the just equality is *ex ante* amounts to saying that the individuals are accountable for the occurrence of this risky event (see developments in Kolm 1998b, and, more briefly, 1996 chap.14). But this accountability cannot be justified by responsibility since the concept of responsibility requires free choice and the individual actually does not choose to take this original risk.⁴¹

9. Moral time-sharing and the multiple self

However, theories of the original position propose solutions to the major problem *P*: *derive the social ethical principle of distribution from society members’ own views while their interests are opposed* (it in fact solely uses their self-interests), *in assuming that the individuals become as concerned with others’ fate as they are with their own, and given that they have only one self*. The other ways of solving this problem are few and undeveloped, and they may face problems similar to those of the theories of the original position (with, also, similar solutions). Then, one may retreat to second-best solutions, that is, to facing problems which give up some requirement of problem *P*, while remaining close to it. The methods for facing problem *P*, apart from the original position, are moral time-sharing and the multiple self.

⁴⁰ See my “Chance and Justice” (1998b) The remark of this section of course adds to all the other foregoing issues, notably that of the multiplicity of views in the original position and of the ways of dealing with it, and can be made precise in these contexts.

⁴¹ This issue of equality in the original position is noted by Harsanyi (1953). He proposes to discard it – as for the particular emotions raised by uncertainty and risk. However, it is related to the very logic of justice and hence of the issue of attributing an ethical value to the concept of an original position. It is remarkable that Harsanyi’s seminal short paper was a reaction to Friedman’s and Savage’s (1952) assertion that VNM cardinal utility is not the utility to be used for social welfare, which implies, given the implicit “welfarism” of their remark, a denial of the assimilation of social ethics and justice to individual choice in uncertainty. Let us add that, in Harsanyi’s 1953 paper, the notional risk is to have individuals’ incomes rather than their utility, and hence this is only a theory of the “partially original position” of the type of “self ascription” (Kolm, 1985, 1998b). The full “thin veil” appears only in the works of 1976 and 1977 (Mongin 2001 well describes this evolution).

Moral time-sharing consists of considering that one individual is successively each of the actual individuals for equal durations and without time discounting (one can for instance assume that the total duration of a sequence is sufficiently short, possibly with repetition of the sequence, but one may also consider time discounting with a corresponding differentiation of the duration of each embodiment). Then, transfers of distributive justice correspond to the individual's saving and borrowing. The corresponding rule of justice can be estimated from these savings and borrowings, as it can be from insurance and risk premia in the case of a theory of the original position, in using observed behaviour, statistics, questionnaires, or introspection. These actual estimations can use simplified models, with samples or the assumption that the individual belongs to types of persons or is in types of situations successively in time or as uncertain possibilities.

Another alternative solution is the consideration of an individual with a *multiple self*, each subself being an actual individual. However, the psychological theories of the multiple self either consider successive selves in time and hence amount to the previous solution, or consider only a couple of selves with very different roles (Freud's triplet of the *id*, the *superego* and the *ego* is an example, though, in fact, the *ego* is the synthesizer between the demands of the *id* and of the *superego* and the constraints called the "principle of reality").⁴²

In any event, these two ways of solution will a priori meet basic problems of the same type as those of the theory of the original position. First, we need a specific individual's preferences for allocating resources among one's successive needs in moral time-sharing, and among one's various drives with a multiple self – as we needed, in the original position, one individual's evaluation of the uncertain prospect. And individuals' preferences again have two reasons to differ in this respect. They may value differently being different individuals, except if there is a fundamental utility. And they a priori value differently the time profile of their instantaneous satisfaction or situation – for the case of time-sharing. Solutions to this problem of multiplicity can be proposed, however, as with the theory of the original position. Second, these ways of solution should be compared with the proper social ethical choice, notably as concerns distributive justice. With moral time-sharing, equality amounts to the individual's evening out the fluctuations of her satisfaction in time. The same problem as that emphasized in the case of risk is present. Some persons may prefer a high and intense enjoyment at some times at the cost of relative insatisfaction at others. But, in the distributive analogy, the individuals who are imposed a poor situation may be unjustly treated. Hence, for the same reason as with uncertainty, moral time-sharing will tend to underscore the ethical value of higher equality: fluctuation-aversion will tend to be lower than the proper inequality-aversion. On the other hand, the "separability" has much less reason for allocation in time than as concerns uncertainty and justice. Indeed, interdependences can be introduced by the effects of habit or boredom (getting used to something or tired of it), by memory and anticipations, and recollections of anticipations and anticipations of recollections. In the interpersonal case, this translates as externalities, and moral time-sharing may be better suited as an analogue for questions of justice among people related by their concern for others due to their own sense of justice or altruism, or to imitation or quest for distinction.

10. Ethical social choice from sociology, anthropology and information.

⁴² A formal model of this theory, and suggestions for its use in social choice, can be found in my book *Pluridimensional Man* (1987).

10.1 Summary of the logic of the question

The foregoing full-concern endogenous solution to social choice results from three requirements: the basic ones of *endogeneity* – derive the solution from society members' views only – and of *impartiality* – a necessary property of justice – can be met with the most demanding requirement of *full-concern* – each member should see each other as herself and care about her as she cares about herself. This raises three conceptual questions: the question of “*transmigration*” – that is, the meaning of “being” someone else and feeling as she does –; the question of *comparison* – how to compare and share among these various embodiments' interests or values –; and, to begin with, the question of *plurality* – how can one single self be several. Three conceptual solutions are considered: the *original position* using uncertainty, *moral time-sharing* using time, and the *multiple self*; and the former is analyzed in detail. These solutions, however, meet two problems: the *multiplicity* of individuals' evaluations, and the *reduction* of justice to a type of self-interested choice (in uncertainty for the original position). Three solutions to the problem of multiplicity are proposed: *regress*, which can be seen as favored by consistency, and *aggregation* or a *social contract* among impartialized views.

There remains the issue of reduction to self-interest. This problem can be avoided, while keeping endogeneity, in retreating from full concern, and in considering that individuals consider all individuals (including themselves) as being in the same type of social relations with them. The reduction problem is avoided because such less-than-full concern does not face the questions raised by the full-concern solutions: It avoids the question of the plurality of selves and proposes a solution for that of interpersonal comparison, while it replaces the “transmigration” issue by the *distanciation* or *objectivization* of seeing oneself as only one other. Its impartiality results from a notional *homogeneization* or *uniformization* of the social relations by the *transformation* of those which are not the reference one into the latter.

However, the less-than-full-concern endogenous solutions have to choose the social relation of reference, and they also meet the problem of multiplicity (as the full-concern solutions do): there will a priori be one evaluation by individual. The differences among these evaluations, however, are much reduced, compared to those between the actual self-centered views, by the impartialization. These differences are also limited by social norms. And they can be erased by the general methods which can be used for moral opinions, the mutual full information about the causes of others' views and about the sentiments they raise (this issue will be developed in section 10.3).

10.2 Extension of close social concern

The less-than-full-concern endogenous solutions derive the norm of distributive justice from the way in which society's members actually compare and distribute among others who equally concern them, and among whom they feel an equal obligation of impartiality and fairness. As noted in section 2, such a reference type of social relation is chosen, and each member's judgment is “impartialized” in having her see all members (herself included) as if they were in this type of relation with her – which also implies a type of relation among themselves. We thus consider members' views if all members were, for instance, her brothers (fraternity), her children (a kind of paternalism), her first degree cousins one sees once a year (say), her nephews (impartial extended “nepotism”), her neighbours, or common members of the society.

This impartialization makes individuals' views much closer to one another than they were in the self-centered reality. Moreover, a common culture and social norms may render these evaluations similar for most members. This would provide the ethical norm with an anthropological input, which may be a general necessity. However, morally valid norms should sufficiently respect individuals' autonomy of choice with regard to social pressure or lack of individual reasoned consideration.

These solutions raise a number of issues. The differences between the impartialized and actual views of individuals may elicit social and political limits to the possible implementation of the rule of justice. We have seen, however, that people may in fact agree with this coercion at a deeper level of consideration. Indeed, the general situation of justice and particular transfers have an aspect of public goods (collective concern). Then, coercive implementation checks free riding and guarantees each individual that others contribute their fair share. Moreover, moral coercion can be approved by individuals' moral selves and internal "impartial spectators", for checking their selfish self's misbehavior, given the common weakness of their will in this respect, that is, as a remedy to moral *akrasia*.

Closer social relations of reference imply a particular care for individuals' needs and desires and may be more moral. Then, for less than full concern, justice can be the extension, to the larger society, of rules of concern, sharing, mutual aid and solidarity that naturally exist among siblings or other close relatives. Brotherhood, or the republican ideal of fraternity, can be the ideal of justice. However, in actual relations of this type the allocation is a consequence of the interpersonal sentiments which also have a major social value in themselves. Hence, the next social improvement, after justice, consists of steps in the direction of all society members having sentiments of this kind among themselves. A first step in this way can be found in justice taking its ancient sense, that is, an individual's attitude and behaviour toward others rather than only abstract principles. The progress in social sentiments can be promoted by culture, education, suasion, imitation, trust, and a deeper social consciousness. The inner moral self should be more positive toward others and involved in their favor than a "spectator" is. In this measure, justice can rest on voluntary implementation and hence be reconciled with freedom. This constitutes progress in the direction of a society of reciprocity⁴³. As Aristotles says: "If you have friends, you do not need justice, but you have justice, you want friends in addition".

10.3 Homogeneization through causal, formative and empathetic information

However, some problem of multiplicity may remain. Individuals may compare or share differently among people in the reference relation with them. Morally acceptable social norms may not suffice to homogeneize their views or behavior in this respect. The solution, then, consists of applying solutions more generally applicable to all moral and social views and which rely on concepts of information. This solution can be applied to actual views or to views already more or less "moralized" such as the notionally impartial evaluations as concerns justice. The crucial issues will refer to individuals' information, though this concept will be applied beyond its usual applications and understandings. Equal information of the individuals will in particular be considered. This can result from full information of all. This fullness can be, and will often be, notional, hypothetical. But full information is a natural condition of rationality for the present objective of constructing a pure, general ethical theory:

⁴³ This is the topic of my book on *General Reciprocity* (Kolm 1984).

this construct can hardly be based on ignorance, misunderstanding, confusion, and mistaken beliefs.

Equal and full information will erase most, and, in the end, all, differences in individuals' correspondingly impartialized views. This will first affect standard, ordinary, relevant information. This will also affect the knowledge of the relevant arguments, reasonings, and actual positions. Moreover, an individual's social or ethical view depends on the information and views about facts and values she has been exposed to in her social environment, in particular her education. It can also depend on her various life experiences, and the knowledge of these experiences and of the feeling they elicit can be conceived of as information. Hence, these causes of her views can be seen as "formative information". If all the foregoing information were identical for all individuals (notably each fully knows all the cases), the only possible cause of differences in individuals' views would be differences in innate sensitivity.

Moreover, individuals' information includes knowing others' views. This implies knowing what these views are and their reasons and causes. But this also implies knowing how it feels to have this judgment, an "empathetic information". And full empathetic information implies experiencing this feeling with the same intensity, duration and involvement. Moreover, this information should be for all, hence mutual between all pairs of individuals, with all the implied interdependences. And individuals cannot jointly have incompatible sentiments and opinions. The outcome can only be a kind of fuse of minds and hearts. Such states sometimes exist. At any rate, this can constitute a theoretical focal point for the solution.

This conception can be given more realism in considering it as the state of convergence of a process of increasing individual information, notably about the reasons and causes of others' views and about the sentiments associated with them. This can notably result from interpersonal dialog, where information, reasons, arguments, and also affects, emotions and sentiments, are transmitted and exchanged. People learn, inform, convince and are convinced. This process is endogenous to the considered society. It can be actual, theoretical, or a mix of both.⁴⁴ Note that the reference to the considered full information is essential (agreement per se cannot be a criterion of optimality).

11. Conclusion

To conclude, let us first note that clarifying the theory of the original position seems to be very necessary for the clarity and rationality of the present social ethical discourse and debate: Rawls constructed such a theory for justifying the replacement of utilitarianism by something else, while an original position is the main justification of a utilitarian form for Harsanyi and his followers. In any event, this theory proposes a solution for solving the very important problem of deriving a social norm from the views of society's members with opposed interests, in assuming that they come to care for each other as much as for themselves, and given that they only have one self. However, the most rational use of a concept of an original position leads neither to Rawls's principles nor to one utilitarian-like structure of the type proposed by Harsanyi, but to one such utilitarian form for each individual. This is so because

⁴⁴ Models of this process are presented in Kolm 1984, 2000. This of course relates to a very long tradition of actual practices in many cultures and considerations in political science and philosophy (deliberative democracy, "discourse ethics", views of Dewey, Perelman, Apel, Habermas, and so on).

an individual in the original position should at least use her own preferences concerning uncertainty, and the purpose of the theory requires that the uncertainty only be about being the actual individuals (a “thin veil of ignorance”). However, the views of the individuals in the original position will be much more similar to one another than those of the actual individuals. Moreover, the problem of the multiplicity of views in the original position is of the same nature as that of the basic problem to be solved – the actual differences in individuals’ views –, and hence consistency suggests, and probably demands, that it be solved the same way. This leads to a regress of original positions converging to a unique, unanimous, evaluation. Other possible solutions consist of an “original position aggregation” and in resorting to a genuine original social contract.

The results only depend on the fact that “being someone else” implies having her preferences about the actual state of the world. However, they are more specific and simpler in the cases where the individuals agree about the desirability of being the various individuals (given the state of the world). This structure avoids types of inconsistency which can be entailed by preferences about being someone else, and it can result from a common social conception of happiness or the good life (fundamental utility).

However, an individual choice in uncertainty a priori differs from a choice of justice among individuals. This shows as concerns possible reasons for giving more to some at the cost of less for others, the choice of the relevant items, and the phenomenology of choice. Other solutions to the noted problem, such as moral time-sharing, are bound to raise similar difficulties. One would then have to resort to social ethics defined as extending to a larger society the ways of sharing and mutual aid standardly used among individuals in a closer relation. Remaining differences in individuals’ (impartialized) views are erased by the notional homogenization of the “formative information” that formed people’s views and the consideration of people’s “empathetic information” about others’ moral sentiments.

This general approach constitutes the endogenous way of solution to the problem of social choice. It calls just and good what society calls this, because of epistemic necessity (where else can the answer be?), an attitude of respect, and non-arbitrariness. In this sense, it is naturalistic (it derives ought from is). “Society’s view”, however, is its members’ unanimous informed view, since endogeneity and respect forbid opposing it, and this condition suffices for the solution. Unanimity suffices first of all because views about justice have to be impartial, hence ban self-centeredness and exclusive self-interest. Such “impartial spectators” exist in many people who occasionally reveal them, and they can be built from individuals’ overall actual views. The moral differences that may remain are erased by successive impartial views of impartial views (regress), by a social contract agreement among these moral views or an aggregation of them, and by individuals’ full “formative information” about causes of moral judgments and “empathetic information” about moral sentiments.

References

- Amiel, Y. and F. Cowell. 1999. *Thinking about Inequality : Personal Judgment and Income Distribution*. Cambridge : Cambridge University Press.
- Archibald, G.C., and Donaldson, D. 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.
- Barrow, R. 1980. *Happiness*. Oxford : Martin Robertson.
- Boudon, R. 1999. Vox populi, vox dei? Le “spectateur impartial” et la théorie des opinions. Colloque sur l’explication des normes sociales (Paris, Octobre 1999).
- Edgeworth, Y. 1888. *Mathematical Psychics*.
- Frankel Paul, E.F., F. Miller, and J. Paul ; eds. 1999. *Human Flourishing*. Cambridge : Cambridge University Press.
- Friedman, M. and L.J. Savage. 1952. The expected utility hypothesis and the measurability of utility. *Journal of Political Economy* 60,463-74.
- 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* 61,434-435.
- Harsanyi, J. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63,309-321.
- Harsanyi, J. 1976. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: P. Reidl.
- Harsanyi, J. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
- Kolm, S.-Ch. 1963. *Les Fondements de l’Economie Publique (Introduction à la théorie du rôle économique de l’Etat)*. Paris : IFP.
- 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, pp.109-77, and *Public Economics*, London, Macmillan, 1969, pp.145-200.
- Kolm, S.-Ch. 1971. *Justice et Équité*. Paris : Cepremap (Reprinted 1972, Paris, CNRS).
- Kolm, S.-Ch. 1984. *La Bonne Economie, La Réciprocité Générale*. Paris. PUF.
- Kolm, S.-Ch. 1985. *Le Contrat Social Liberal*. Paris : Presses Universitaires de France.
- Kolm, S.-Ch. 1987. *L’Homme pluridimensionnel*. Paris : editions Albin Michel.

- Kolm, S.-Ch. 1990. The normative economics of unanimity and equality: equity, adequacy and fundamental dominance. In K.J. Arrow, ed. *Markets and Welfare*. London: Macmillan, pp.243-86.
- Kolm, S.-Ch. 1993. *Equal Liberty*. Paris: CGPC.
- Kolm, S.-Ch. 1995. The economics of social sentiments: the case of envy. *Japanese Economic Review*. 46 : 63-87.
- Kolm, S.-Ch. 1996. *Modern Theories of Justice*. Cambridge, MA : MIT Press.
- Kolm, S.-Ch. 1998a. *Justice and Equity*. English translation with new foreword. Cambridge, MA : MIT Press.
- Kolm, S.-Ch. 1998b. Chance and justice : Social policy and the Harsanyi-Vickrey-Rawls problem. *European Economic Review* 42: 1393-1416.
- Kolm, S.-Ch. 2000a. Introduction, In L.-A. Gérard-Varet et al., eds. *The Economics of Reciprocity, Giving and Altruism*. London : Macmillan.
- Kolm, S.-Ch. 2000b. Freedom Justice. Université de Caen.
- Kolm, S.-Ch. 2001a. Unequal Thoughts. *Journal of Economics* 73(1), 96-106.
- Kolm, S.-Ch. 2001b. *Macrojustice*.
- 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 S. Barberá, P. Hammond and C. Seidl, eds. *Handbook of Utility Theory* Vol.1. Boston : Kluwer Academic Press, pp.373-342.
- Nagel, T. 1986. *The View from Nowhere*. Oxford: Clarendon Press.
- Nagel, T. 1991. *Equality and Partiality*. Oxford: Clarendon Press.
- Pattanaik, P. 1968. Risk, impersonality, and the social welfare function. *Journal of Political Economy* 76,1152-1169.
- 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 A. Sen and B. Williams, eds. *Utilitarianism and beyond*. Cambridge: Cambridge University Press.
- Sen, A. 1970. *Collective Choice and Social Welfare*. Amsterdam: North Holland.
- Smith, A. [1781], 1976. *The Theory of Moral Sentiments*. Indianapolis: Liberty Classics.

- 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* 13,319-333.