

Social Choice and Just Institutions:New Perspectives

Marc Fleurbaey

▶ **To cite this version:**

Marc Fleurbaey. Social Choice and Just Institutions:New Perspectives. 2006. halshs-00121378

HAL Id: halshs-00121378
https://shs.hal.science/halshs-00121378

Preprint submitted on 20 Dec 2006

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

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

Social Choice and Just Institutions:

New Perspectives

Marc Fleurbaey

CNRS-CERSES, Paris, and IDEP

marc.fleurbaey@cerses.cnrs.fr

Jan. 2002

Revised, December 2005

Abstract: It has become accepted that social choice is impossible in absence of interpersonal comparisons of well-being. This view is challenged here. Arrow obtained an impossibility theorem only by making unreasonable demands on social choice functions. With reasonable requirements, one can get very attractive possibilities and derive social preferences on the basis of non-comparable individual preferences. This new approach makes it possible to design optimal second-best institutions inspired by principles of fairness, while traditionally the analysis of optimal second-best institutions was thought to require interpersonal comparisons of well-being. In particular, this approach turns out to be especially suitable for the application of recent philosophical theories of justice formulated in terms of fairness, such as equality of resources.

Introduction

Three main points are made in this paper. First, the aggregation of individual non-comparable preferences into reasonable social preferences is largely possible, in spite of Arrow's impossibility theorem of social choice. Second, this approach to preference aggregation offers new possibilities for the analysis of the design of optimal institutions. Third, Rawlsian theories of justice will find here a valuable tool for concrete applications of fairness principles in the design of just institutions.

In short, the approach described here opens up possibilities that, so far, have been largely ignored. The received view is that the aggregation of individual preferences requires interpersonal comparisons of well-being *based on information that is not contained in non-comparable preferences*. I will show that this view is not correct, and explain that it is due to unreasonable demands made on the aggregation procedure. Once the unreasonable requirements are relaxed, quite attractive aggregation possibilities appear.

Another common view, at least in economics, is that the design of redistributive institutions such as tax schemes and social security systems is best done by maximizing a social welfare function of the traditional kind, embodying interpersonal comparisons of well-being. This view is vindicated by the pessimistic interpretation of social choice that has just been described. Now, if interpersonal comparisons of well-being are no longer needed in social choice, they can be avoided in institution design, too, and this is why new possibilities are automatically created for this field as well.

A striking convergence between this approach and Rawlsian theories of justice is that, in both, reference to interpersonally comparable subjective satisfaction is avoided in the discussion of the fair allocation of resources. In Rawlsian theories of justice,[1] this is based on the thesis that individuals should assume responsibility for their ends and ambitions in life, so

[1] Such as Rawls (1971), Dworkin (2000), van Parijs (1995).

that a low level of satisfaction about one's achievements, in itself, does not give any claim on social help.[2] In the traditional approach to social choice and welfare economics, one also finds a reluctance to handle utility functions, as opposed to ordinal non-comparable preferences. But in this case it mainly comes either from a rather radical (and contentious) "no bridge" objection to the perspective of making interpersonal comparisons of subjective utility,[3] or from the more pragmatic observation that no satisfactory method of measuring utility or satisfaction in an interpersonally comparable way has yet been developed.[4]

The approach which is presented in this paper, hopefully, will therefore appear useful to any economist or philosopher who, for any of these reasons, is looking for a way to aggregate individual ordinal non-comparable preferences and discuss the design of institutions and policies on this basis. In particular, in the perspective of applying recent egalitarian theories of justice, what is proposed in this paper is a closer cooperation between political philosophers, social choice theorists, and public economists, in the "assembly line" of just institutions. The first would provide the fairness principles. The second would formulate axioms embodying the principles, and derive social preferences. The third would find the institutions that are the best according to such social preferences.

Hold it! Social choice is possible

Social choice has been dubbed "the science of the impossible".[5] The root of this is obviously to be sought in Arrow's (1951) impossibility theorem. Although the theorem was not taken

[2] These theories are sometimes misunderstood as implying that people suffering from terrible pains are not eligible to any help! Satisfaction over one's life must not be confused with experience of pleasure and pain. This is discussed in more detail in the last section.

[3] Robbins (1932), in a positivist stance, argued that interpersonal comparisons of utility were necessarily normative and could not be based on empirical observations only. He was understood as barring interpersonal comparisons from economics, although he later (Robbins 1981) clarified that normative economics in general and interpersonal comparisons in particular were part and parcel of policy advice.

[4] A recent branch of the economic literature studies satisfaction surveys and often uses such statistics as if answers (on a scale from 1 to 4 or 7) were cardinal and directly comparable. In this case, the usual excessive shyness of economics about interpersonal comparisons seems to vanish entirely. For reviews, see Oswald (1997) and Frey and Stutzer (2002). A critical discussion is made in Fleurbaey and Hammond (2004).

[5] See e.g. Hammond (1987).

very seriously by the specialists of welfare economics at the time of its publication, its influence has been growing because no good preference aggregation procedure has ever emerged from welfare economics, while the impossibility theorem has turned out to be very robust, and its main content, reproducible in every imaginable context, including economic models of resource allocation. Another reason for this influence is that Arrow's seminal contribution, although negative in its substance, was also very positive in terms of method. The concept of a *social choice function*, mapping various profiles of population preferences into "aggregate" social preference relations, was a beautifully elegant theoretical object, and the idea of relying on the axiomatic method in order to look for a good function was perfectly appropriate. The same remarkable contribution created the theory of social choice… and put it at a deadlock.

Let the theorem be re-examined. The set up is the following. There is a set of options (social states, policies), and a finite population of individuals. Every individual has preferences over the set of options, so that, when considering any pair of options x, y, she can say whether she strictly prefers x, or y, or is indifferent between x and y. Formally speaking, by "preference" or "preference relation", I mean a binary relation "at least as good as", which is a complete preorder (that is, transitive and complete).[6] And one says that x is strictly preferred to y when x is at least as good as y, whereas y is not at least as good as x; indifference between x and y means that any of these two options is at least as good as the other one. A *profile* of population preferences is a list describing every individual's preferences.

In this set up, the social choice problem is to define a "social" preference relation on the same set of options, as a function of the population's preferences. A *social choice function* is a mapping which defines a social preference relation for every profile of the population's

preferences in some domain. Intuitively, the purpose of the social choice function is to make a synthesis of the population preferences. A social preference relation is, formally, the same kind of object as an individual preference relation, namely, a complete preorder over the set of options. But unlike individual preferences which are real characteristics of existing individuals, social preferences are a pure construct, and are meant to provide guidelines for social decisions.[7] They should not be viewed as subjective preferences of some kind of collective entity.

In later formulations of Arrow's theorem, the following four conditions, or axioms, are imposed on the social choice function:

Unrestricted domain: The social choice function must be defined for every conceivable profile of population preferences.

Weak Pareto: When all individuals strictly prefer one option to another, social preferences must follow.

Independence of Irrelevant Alternatives: Social preferences on a pair of options should only depend on the population's preferences on these two options.

Non Dictatorship: There is no individual such that, for all profiles, the strict part of his preferences is always obeyed by social preferences.

The theorem says that these four conditions are incompatible.

Arrow's Impossibility Theorem: *There exists no social choice function satisfying the four conditions, when the set of options contains at least three options, and the (finite) population has at least two individuals.*

---

[6] The preference relation is transitive if whenever option *x* is at least as good as option *y*, and the latter is itself at least as good as option *z*, then *x* is at least as good as *z*; complete if for any pair of options, one of them must be at least as good as the other one.

[7] This distinction should not be overstated. As we will see, individual preferences may also be a construct. And for some moral theories, the morally good ranking of options is also a real thing.

I now proceed to show that the theorem is not so relevant as usually thought, because the set of conditions is unreasonably demanding. Of the four conditions, only the last one is totally unquestionable. No sensible aggregation procedure can ever let one individual always impose his strict preferences. In fact, one could readily require a stronger condition of *anonymity*, that is, require that social preferences do not depend on individuals' names, or, equivalently depend on individual preferences in a symmetrical way (everyone should have an equal “weight”).

The other three axioms can be criticized, but to unequal degrees. Unrestricted Domain is unduly demanding in many contexts in which reasonable individual preferences cannot be any arbitrary preorder over the options. Economists are used, for instance, to work with individual preferences which satisfy particular properties (continuity, convexity, etc.) in addition to being complete preorders. Even in non-economic contexts, such as voting, preferences are usually shaped in special ways (a conservative voter does not rank a communist candidate above a social-democrat). This criticism is not very important for our purposes here, because Arrow's theorem is robust to many domain restrictions, in particular economic ones.

The Weak Pareto criterion can also be criticized, from non-welfarist quarters especially. Satisfaction of individual preferences may not be all that matters in terms of individual well-being. More basically, ordinary preferences cannot be taken at face value, because they may embody irrational or anti-social impulses, framing effects, imperfect information, etc. All such criticisms, however, do not really call for abandoning the Pareto principle, but only for reformulating it in terms of an appropriate notion of well-being or personal good. As a consequence, I would like, from now on, to assume that individual

preferences reflect individual well-being or personal good in an ethically acceptable way, whatever that is.[8]

My main target of criticism will be the axiom of Independence of Irrelevant Alternatives (hereafter IIA). As Hansson wisely wrote, "Arrow's theorem is really a theorem about the independence condition".[9] I do not want to be understood as saying that IIA is an axiom based on unsound principles. It would certainly be a good thing if the social comparison of two options could depend only on individual preferences on these two options and on nothing else. That would make social choice a very simple matter. Very little information would be needed in the aggregation process. But simplicity and informational parsimony are not all that counts. Ethical relevance is also important. And IIA does wipe out ethically relevant information for the social comparison of two options.

As an example of relevant information being wiped out, consider Ann and Bob, who, in the status quo, have the following bundles at their disposal. Ann has ten apples and two oranges, while Bob has three apples and eleven oranges. Would it be a good thing to transfer one apple and one orange from Bob to Ann? Let us focus on the profiles of preferences in which both individuals always prefer having more of each good. In all such profiles, Ann strictly prefers the transfer, whereas Bob strictly prefers the status quo. But IIA requires social preferences to depend only on those pairwise preferences. That is, knowing that Ann prefers the transfer and Bob the status quo should be, on the basis of IIA, enough information for deciding whether social preferences should approve the transfer, prefer the status quo, or be indifferent.

[8] Arguments to the effect that social goodness does not depend only on individual well-being can also be alleviated by a proper adaptation of the well-being or personal goodness concept. Take for instance the view that making everyone better-off but with less knowledge is socially worse. It suffices to incorporate knowledge in the definition of well-being in order to conclude that individual well-being has not really increased in this case. Or consider the example of a sadist whose personal well-being may increase when his sadistic impulses are satisfied whereas social preferences should simply ignore this increase and not balance it against the suffering of others. Again, it suffices to measure individual well-being in a "socially correct" way in order to avoid this kind of difficulty.

[9] Hansson (1973), p. 25.

With these scarce data, I am afraid indifference is the only reasonable option. Indeed, the status quo displays bundles (10,2) for Ann and (3,11) for Bob, and after transfer the bundles become (11,3) for Ann and (2,10) for Bob. There is a perfect symmetry here between goods and between people, which makes it impossible to find any reason to prefer one option over the other. But, and this is my main point, there may be other relevant information that should influence social preferences. Suppose that in a particular profile P, both individuals have identical preferences, and are indifferent between the bundle (10,2) and the bundle (3,11). Then, in this profile, the status quo provides Ann and Bob with bundles which both find equally valuable, whereas the transfer would make Bob unambiguously worse-off than Ann, since both would agree that his bundle would be less valuable, and he would envy Ann in the sense of the envy test of the theory of fairness (i.e., he would rather have Ann's bundle). Preferences in profile P can even be assumed to be such that the status quo is a Pareto-efficient allocation,[10] whereas the transfer would destroy efficiency (if marginal rates of substitution differed between Ann and Bob at the resulting allocation).[11] Reasonable social preferences may certainly prefer the status quo on these grounds: An efficient envy-free allocation, when individuals have identical preferences, is better than an inefficient allocation with envy. Now, in another profile P', the situation might be reversed. In this new profile, both individuals might be indifferent between (11,3) and (2,10), so that the post-transfer allocation would then be the (efficient, envy-free) one to be socially preferred.

In other words, it would be quite sensible for social preferences to rely on such information as "Bob prefers Ann's bundle", or Pareto-efficiency of the allocation. But this is, unjustifiably, forbidden by IIA. Notice that the kind of information that this example shows to be relevant belongs only to non-comparable ordinal preferences.

[10] A Pareto-efficient allocation with all individuals on the same indifference curve is, moreover, also a Walrasian equilibrium with equal budgets (also called sometimes an equal-split competitive auction).

One may be worried that this example is about an economic problem with transferable goods and particular preferences, and imagine that IIA remains valid for the abstract contexts usually studied in social choice theory, in which the options have no special structure so that, for instance, a fairness concept such as envy cannot be applied. It is true that in abstract issues less information about options and thereby about individual preferences is available. But this does not totally alleviate the above objection to IIA. Consider two options, *x* and *y* and two individuals, Alice and Brian. Suppose we know that Alice prefers *x* and Brian prefers *y*. According to IIA, this is enough information to determine the social preferences over *x* and *y*. What can social preferences be on such a poor informational basis? They should probably declare indifference again, as in the previous example. But suppose we now add the information that, with profile P*, in *x* and *y* Alice is at her top and second best options, whereas Brian is at his worst and best options, respectively. Then it might become sensible to prefer *y*. With *y* one individual is at her second best and the other at his best, whereas with *x* one individual would get his worst option. The reverse situation might occur with another profile P**, leading social preferences to prefer *x* then, instead of *y*. Again, this additional information about the ranks of options in individuals' preferences may be deemed relevant by reasonable social preferences, and it seems questionable to exclude it, as IIA does.

Now, the fact that IIA unduly excludes such information does not prove that such information would be sufficient to obtain a possibility instead of the impossibility theorem. The usual view is that it is not, on the ground that the impossibility remains as long interpersonally comparable indices of well-being are missing. I now proceed to show that this view is not justified, by presenting examples of reasonable social choice functions that make use of more information about individual preferences than allowed by IIA, but of no other information than individual ordinal and non-comparable preferences.

[11] One might object that the post-transfer allocation is not Pareto-inferior to the pre-transfer allocation, so that the two distributions appear efficient. But the correct definition of efficiency refers to the possibility of third

Let us consider the abstract context first. This is the least favorable context, because, as explained above, preferences and options are described with minimal information. Nonetheless, there are several interesting social choice functions for this setting, which violate IIA on reasonable grounds, and satisfy other valuable properties. The best known is the Borda rule, which compares options on the basis of their average rank in people's preferences. An option is better if it occupies, on average, a higher rank in individual preferences. This violates IIA because, quite reasonably, it examines whether going from *x* to *y* represents a big or small step in people's preferences, which depends on how many options are in between (for instance, the Borda rule prefers *y* under profile P* and *x* under profile P**, in the above example). A "maximin" variant of the Borda rule, which seems to have been neglected in the literature, compares options on the basis of their worst rank in people's preferences. This still violates IIA because it compares options on the basis of how many options are still considered worse than them. In the above example, the worst rank of *y* in P* is 2 (corresponding to Alice's preferences), the worst rank of *x* is equal to the number of options (in Brian's preferences), entailing a clear social preference for *y*. Another important rule is the Kemeny-Young rule, which looks for a complete preorder that minimizes the total number of disagreements with the majority rule on pairwise comparisons. It has been axiomatically justified by Peyton Young,[12] and has recently received a vigorous philosophical defense by Risse (2001). This rule violates IIA because it discounts individual preferences which are outlandish, in the sense that they too often contradict the majority verdict in pairwise contests. It then often requires extensive information about all majority comparisons before judging on two particular options.

Notice that, in the case where there are only two options in the set of options, the three rules coincide with the majority rule. In summary, these three examples are quite reasonable

---

allocations. A distribution is inefficient if one can redistribute its resources so as to make everyone better-off.

[12] See esp. Young and Levenglick (1978).

extensions of the majority rule, based on defensible violations of IIA. But the truth is that no revolution in welfare economics will ever come from these examples. The main reason for this is that the majority rule is a very poor device in distributive justice issues. Following the majority rule in questions of resource allocation is quite unwise because selfish majorities may too easily sacrifice minorities and deprive them of the little they have. This kind of shortcoming is due to the impoverished information available in the abstract framework which is suitable to applications of simple voting rules such as the majority rule. This is why economic issues have to be addressed with richer conceptual tools capturing the fairness issues, and cannot be subsumed under the political model of voting.

Let us now turn to an economic issue, such as the problem of distributing a fixed quantity of divisible goods to a finite population of individuals, and illustrate how relaxing IIA makes it possible, in this context, to find reasonable social choice functions without any other information than ordinal non-comparable individual preferences. François Maniquet and I (1996, 2001) have studied two particular kinds of social preferences for this particular case. The first one compares allocations by focusing on the worst share of the total available resources that an individual would accept to receive instead of her current bundle. For instance, if one allocation has one individual accepting two percent of the total resources in exchange for his bundle at this allocation, whereas in another allocation no one would accept less than three percent, then the latter is to be preferred. Such social preferences bear a close relationship with the concept of fairness as “egalitarian-equivalence” proposed by Elisha Pazner and David Schmeidler (1978), because the best allocations among all feasible allocations will be precisely the “egalitarian-equivalent” ones, in which everyone is indifferent between his current bundle and a given percentage (the same for all individuals) of total resources.

The second kind of social preferences is very similar, except that the shares of total resources are not considered for direct consumption, but for further exchange on the market, at some relevant prices. I will not go into the details of the definition here, but only mention that the best allocations for this second kind of social preferences are the competitive equilibria from equal-split, a famous concept of solution in the theory of fairness.

These two examples violate IIA, because they rely on much more information than pairwise preferences over two allocations. They depend on the shape of large parts of individual indifference curves.[13] For instance, if the total amount of apples and oranges to be distributed between Ann and Bob is thirteen units of each good, "egalitarian-equivalent" social preferences will, if Ann receives the (10,2) bundle, seek the value of X such that she is indifferent between (10,2) and (X,X). But, interestingly, for any two allocations, it is enough for such social preferences to know the individual indifference curves at the contemplated allocations only, in order to be able to rank those allocations. This is a weakened form of IIA, which can be related to a weak independence axiom proposed, independently, by Hansson (1973) and by Pazner (1979).

Weak IIA: Social preferences on a pair of options should only depend on the population's preferences on these two options and on what options are indifferent to each of these two options for each individual.

Is it reasonable to rely on indifference curves in this way, instead of just pairwise comparisons? Yes, and this is precisely what the Ann-Bob example was meant to show. In that example, knowing the individuals' indifference curves at two allocations makes it possible to see, for instance, whether, in one allocation, both individuals are at the same

[13] Arrow's theorem remains valid when IIA is weakened so as to permit the use of information about local parts of indifference curves, such as marginal rates of substitution or even more (see Inada 1964 and Fleurbaey, Suzumura and Tadenuma 2005a). This can be understood intuitively as coming from the fact that local information about preferences (i.e. information about preferences on nearby bundles only) is not sufficient to have a comprehensive picture of an individual's valuation of his bundle.

indifference curve (implying in particular that they are indifferent between their respective bundles), and whether, in the other allocation, one individual envies the other. It also makes it possible to ascertain whether an allocation is inefficient in the sense that it would be possible to redistribute its resources in order to make everyone better-off in the population.

This kind of possibility opens a new field of axiomatic research. Indeed, once IIA is weakened to Weak IIA, it is possible to strengthen Arrow's other conditions without introducing an impossibility. Arrow's axioms of Weak Pareto and Non-Dictatorship are quite minimal. In particular, the latter can be replaced by much more stringent equity requirements. As an example, consider the following requirements.

Transfer Principle for Equals: If two individuals have identical preferences and one has more of each good, then it is socially as least as good to make a transfer of goods from him to the other that leaves him still with more.

Minimal Egalitarianism on Shares: If all agents receive proportional shares of the total resources, it is socially at least as good to replace this by an equal split of this allocation.

These are just examples of equity conditions, among many others which may be imagined. In another paper,[14] I have shown that the Pazner-Schmeidler kind of social preferences defined above is singled out by the combination of these two axioms with Weak Pareto and Weak IIA. A noticeable feature of this result is that such social preferences are of the maximin kind, since they focus on the worst-off individual, as measured by the share of total resources accepted, whereas none of the four axioms involved in this result does, by itself, imply such absolute priority for the worst-off. In particular, Transfer Principle for Equals is only about transfers in which the recipient gets exactly what the donor gives, and Minimal Egalitarianism on Shares is only about comparing perfect equality to inequality.

[14] See Fleurbaey (2005), Th. 2.

None of these two axioms says anything about cases when the worst-off are helped at a great cost to the better-off.

This result is only an example and there exist other axiomatic justifications for these kinds of social preferences.[15] The lesson which may be drawn here is that social choice is definitely possible in the absence of interpersonally comparable utility functions. A reasonable weakening of IIA appears amply sufficient to give way to the construction of equitable and Paretian social preferences.

The rest of this section is devoted to a brief discussion of two other independence requirements. The Pazner-Schmeidler social preferences introduced here not only satisfy Weak IIA, but also make the evaluation of feasible allocations (i.e. allocations with a total consumption less than or equal to the available quantity of each good) independent of individual preferences over bundles which contain more of some good than the available quantity. This may be shortly described as meaning that *the ranking of feasible allocations does not depend on individual preferences over "infeasible allocations"*.[16] Of course, if the set of actually attainable allocations, in a given context, is only a strict subset of this feasible set,[17] then the ranking of attainable allocations by Pazner-Schmeidler preferences will typically depend on individual preferences over non-attainable (though feasible) allocations. There appears to be nothing wrong in general, from an ethical viewpoint, with using information about infeasible or unattainable alternatives in order to evaluate individual situations and rank social alternatives.[18]

---

[15] See e.g. Fleurbaey and Maniquet (2001), Fleurbaey (2004, 2005).

[16] One then sees that the Pazner-Schmeidler social preferences satisfy a combined independence condition formulated as follows : *Social preferences on a pair of feasible options should only depend on the population's preferences on these two options and on what feasible options are indifferent to each of these two options for each individua*l.

[17] This may happen for instance when individuals have initial endowments and are willing to hide or destroy them when threatened by redistribution. Then some allocations which are "technically" feasible are not attainable.

[18] Plott (1976) has introduced an axiom of Independence of Infeasible Alternatives in a related context, and argued that it is necessary for the social criterion to be implementable (a pragmatic rather than ethical argument). This restriction, however, does not apply to second-best applications described later in this paper. (In such

The fact that Pazner-Schmeidler social preferences are independent of preferences over infeasible alternatives, however, comes at a cost. They do depend, in their definition, on the quantity of available resources since individual bundles are evaluated in terms of "equivalent shares" of this total quantity. The axiom of Minimal Egalitarianism on Shares indeed refers to this quantity as well. A priori, it would seem nice to be able to define social preferences as a pure aggregation of individual preferences, without referring to feasibility constraints and then marking a clear separation between objectives and constraints. This is easily done, for instance by defining modified Pazner-Schmeidler social preferences with a *fixed* reference bundle that serves to compute equivalent shares, independently of the available resources (which may or may not coincide with the reference bundle). Such social preferences, however, will no longer be independent of individual preferences over infeasible allocations. There is an impossibility theorem here.[19] But, as we have seen, the idea that social preferences should be independent of individual preferences over infeasible alternatives is not that compelling. Moreover, the idea that feasibility constraints should not influence the definition of social preferences is not compelling either. It is deceptively tempting to imagine that aggregating individual preferences is a purely normative exercise supposed to reflect the orientation of individual preferences, ignoring all feasibility considerations. This idea is attractive in the abstract only. In concrete examples, knowing the relative scarcity of available resources is important and points to a particular way of evaluating bundles. We are used to doing that in daily practice, for instance, when relying on current market prices for the evaluation of consumption bundles.

---

applications, it is assumed that the decision-maker knows the statistical distribution of preferences in the population.)

[19] It is impossible to satisfy Weak Pareto, Non-Dictatorship, Independence of Infeasible Alternatives and Independence of the Feasible Set. This is actually a conjecture, since a full proof has not yet been published. For a precise statement of these two independence conditions, see Fleurbaey (2003), which discusses a version of the results involving Pareto Indifference. A slightly weaker result (involving Weak IIA in addition) is proved in Fleurbaey, Suzumura and Tadenuma (2005a).

In summary, the social preferences described here generally satisfy Weak IIA, but not IIA which is definitely too demanding. Some of them are independent of individual preferences over infeasible allocations, others are independent of feasibility constraints, while none of them satisfies both requirements. Fortunately neither of these two requirements appears really compelling.

Be fair with the fairness approach

The title of Amartya Sen's beautiful Nobel lecture is "The possibility of social choice".[20] Sen, along many others, has spent years defending social choice against what they called the "Arrow gloom". Their efforts have, however, concentrated on a different approach than the one highlighted above. They have accepted the thesis that social choice is impossible in absence of interpersonally comparable indices of well-being. But, biting the bullet, they defended the additional, reasonable thesis that interpersonal comparisons of well-being are not so hard to perform. It is not very difficult, in particular, to determine who should be given priority in societies with large inequalities.

Their second thesis is quite acceptable, but it does not establish the first thesis. The examples of the previous section have shown that interesting social preferences can be based on ordinal non-comparable individual preferences. Moreover, it will be argued later that this alternative approach may help to define social priorities in more convenient and more relevant ways. In this section, I would like to clarify a few theoretical points about the differences and similarities between these two approaches. For simplicity of reference, I will call Sen's approach the "comparability" approach, and the alternative one the "fairness" approach (in view of its close connection to the economic theory of fairness,[21] which analyzes the allocation of multi-dimensional resources and defines fairness and efficiency concepts on the

[20] See Sen (1999).
[21] For a survey, see e.g. Moulin and Thomson (1997).

basis of ordinal and non-comparable individual preferences).[22] Similarly, for clarity's sake the expression “fairness concept” in the sequel will refer only to distributive notions which can be formulated on the sole basis of ordinal non-comparable individual preferences.[23]

An important similarity between the comparability and fairness approaches is that both reject IIA. But the essential difference is that they reject different *parts* of IIA. In order to understand this point, it is necessary to change the framework of the social choice problem in order to have not only individual preferences but also interpersonally comparable indices of utility from which one seeks to derive social preferences. Utility need not be understood here in the traditional welfarist sense, and, like preferences above, can represent any concept of well-being that is judged relevant for the evaluation of individual situations in the social choice context, such as, say, a capability index. Suppose, then, that we still have a set of options to rank, and that individuals are now described not by their preferences but by their utility measured on a numerical scale at the various options. This new setting can be viewed as containing more information, since from an individual's utility function, one can retrieve an individual preference relation by defining an option x to be at least as good as another option y for some individual whenever her utility at x is greater than or equal to her utility at y. In what follows individual “preferences” refer to this preference relation derived from the utility function.

In this new setting, one may still look for a social preference relation as a function of the population profile of utility functions. Arrow's axioms can still be formulated, with little change. The Unrestricted Domain now says that all individual utility functions are admitted. The other axioms do not even need to be rephrased. And Arrow's theorem is still valid.

[22] This choice of terminology is not meant to suggest that the comparability approach ignores issues of fairness in a general sense of this word.
[23] The reader who is less interested in theoretical underpinnings than in consequences and applications may skip this section.

Now, in this new formulation, it is important to notice that IIA says two things. First, it says that the comparison of two options should depend only on people's *preferences*, as opposed to the numerical values of utility. Second, it adds that the comparison must rely only on *pairwise* preferences about the two contemplated options. It is even possible to formulate this second restriction in a way that does not presuppose the first one: IIA requires the comparison to rely only on the numerical levels of utility at the two options. Let these two parts be formulated as separate axioms:

Ordinal Non-Comparability: Social preferences should remain the same when the profile of individual utility functions changes without altering individual preferences.

Independence of Irrelevant Utilities: Social preferences on two options should only depend on individual utilities at these two options.

Sen and others have very well analyzed this dual content of IIA. And they have argued that Ordinal Non-Comparability must be abandoned, whereas, in their approach, Independence of Irrelevant Utilities may be kept. This opens up possibilities for social welfare functions which compute social welfare for a particular alternative as a function of individual levels of utility at this precise alternative. In this way the second axiom is satisfied, since the comparison of two alternatives will depend only on the computation of social welfare at those two alternatives. Examples include the utilitarian social welfare function (focusing on the sum of utilities) and the maximin (focusing on the minimum of utilities).

The approach I present here does exactly the opposite. It retains Ordinal Non-Comparability, which implies that interpersonally comparable utility will be ignored, and non-comparable ordinal preferences will be exclusively considered. But it drops Independence of Irrelevant Utilities, and allows the social ranking of two options to depend on features of utilities at other options. This difference may be illustrated with the Ann-Bob example. Recall that the status quo has Ann with bundle (10,2) and Bob with (3,11). Suppose that Ann's and

Bob's utility levels at the status quo are respectively 4 and 5. And that the transfer of one unit of each good from Bob to Ann would reverse these figures. According to Independence of Irrelevant Utilities, this is enough information to make a social decision. Must we prefer the distribution of utility of 4 and 5 to the distribution with 5 and 4? Both the utilitarian and the maximin social welfare function, for example, give equal consideration to Ann and Bob and are therefore indifferent between the two options. They are therefore unable to take account of the fact that, maybe, the status quo is an efficient allocation in which both agents are indifferent between the two bundles, whereas the transfer would destroy efficiency[24] and would make Bob have a bundle that is then unambiguously worse than Ann's. Such fairness considerations require more information about preferences than allowed by Independence of Irrelevant Utilities. If one wants to take account of such fairness criteria, one has to drop this independence axiom. Once this is done, there is no logical necessity to drop Ordinal Non-Comparability any more in order to obtain possibility results. As shown in the previous section, non-comparable preferences are then enough to get possibility results. The received view that social choice requires interpersonally comparable utilities appears valid *only if Independence of Irrelevant Utilities is imposed.*

Another interesting fact, related to Independence of Irrelevant Utilities, obtains when one combines it with Pareto-Indifference, a condition saying that social preferences should be indifferent between two options when all individuals are indifferent, too. Under the Unrestricted Domain axiom, combining the two conditions (Pareto-Indifference and Independence of Irrelevant Utilities) yields the "welfarism lemma", due to Claude d'Aspremont and Louis Gevers (1977). The lemma says that options will be socially ranked

[24] Recall that the definition of efficiency refers to the possibility of the resources of an allocation so as to make everyone better-off. In our example, it may be that with profile P the distribution (4,5) is not dominated by any other feasible distribution, whereas (5.5,4.5) is a possible distribution that dominates (5,4) and proves that the latter is inefficient. Note that both the utilitarian and the maximin social welfare functions satisfy the Weak Pareto principle and do record that (5.5,4.5) is better than (5,4). But one cannot decide if a given allocation is Pareto-efficient by only looking at the utility levels in this allocation.

on the basis of a unique ranking of distributions of utility levels, which can be called the "welfare ordering" and which applies to utility distributions in the same way, whatever the population profile. The utilitarian or the maximin social welfare functions, for instance, always rely on the *sum* or the *minimum* of utilities, whatever the population profile of utility functions. Comparing the sum or comparing the minimum defines their respective welfare orderings. This may seem unexceptional because this is the usual way of proceeding. But a priori it might be possible to have a social choice function applying the sum ranking for certain profiles of utility functions and the maximin ranking for other profiles. The welfarism lemma excludes such complex possibilities, and ensures that if, say, the sum of utilities is the criterion for one profile, then it is always the criterion, for every profile.

It is interesting to note that such "welfarism", in the lemma, is built up differently by the two conditions involved. In and of itself, Pareto-Indifference prevents social preferences to rely on non-utility features of options in the sense that, in a given profile, social preferences can be expressed by a welfare ordering.[25] But this welfare ordering can be specific to the profile and this leaves a lot of possibilities on how to use utility information. The examples of social preferences presented in the previous section, for the distribution problem, do satisfy Pareto-Indifference, but use utility information in a sophisticated way that makes it possible to consider quite elaborate fairness conditions. Now, Independence of Irrelevant Utilities imposes that, for every pair of options, there exists a welfare ordering for that pair, independently of the profile. This is exactly what the axiom itself says. But the welfare ordering can a priori be specific to the pair, it can be the utilitarian ordering for one pair, the

[25] In order to see this, consider two options yielding the same distribution of utilities for a given profile of utility functions. By Pareto-Indifference, social preferences must be indifferent between them, implying that these two options will ranked similarly with respect to any other option. Therefore social preferences are just ranking utility distributions (in the particular profile under consideration).

maximin for another, etc.[26] Combining the two axioms transparently implies that the welfare ordering must be the same for all pairs and all profiles.

When criticizing "welfarism", Sen[27] has targeted both Pareto-Indifference and Independence of Irrelevant Utilities, by advocating that non-utility features may be taken into account so as to violate both conditions. But, deep down, his "non-welfarist" approach merely asks for a proper definition of utility. Once all relevant features of individual well-being have been put in the measure of utility, that is, for instance, when utility is redefined in terms of capabilities or functionings, then Sen is willing to accept both Pareto-Indifference and Independence of Irrelevant Utilities, coming back, formally speaking, to the "welfarism" of the lemma. In contrast, by dropping Independence of Irrelevant Utilities, the "fairness" approach to social choice does no longer satisfy the welfarism lemma. It does not merely apply a welfare ordering to exogenously defined interpersonally comparable utilities.

Nonetheless, it is clear that the fairness approach also performs comparisons between individuals.[28] For instance, the Pazner-Schmeidler example described above compares the percentages of total resources that individuals would accept in exchange for their bundles. Moreover, it can be described as the application of the maximin criterion to such measures of individual "utility". Does not this blur the difference between the two approaches?

There are two essential differences between the ways the two approaches perform comparisons between individuals. The first one is about the general kind of object that is compared, reflecting potentially important philosophical divergences. In the "comparability" approach, comparison is about utility, whatever that is. In the "fairness" approach, the comparison is about bundles, as they are valued by individual preferences, with reference to equity principles. If utility, in the "comparability" approach, is defined in such a way that it

[26] To be rigorously precise, this holds true for every *ordered* pair, because the social ranking on well-being may not be neutral. For instance, it may say that x is at least as good as y if and only if the sum of well-beings is at least twice in x than in y.

[27] Notably in Sen (1979).

merely reflects the value of bundles as assessed through individual preferences, then the two approaches, in their results, come close to each other. If, on the contrary, a more comprehensive notion of well-being is invoked in the comparability approach, then the two approaches lead to social preferences that are substantially different.

The second difference is about the method by which what is compared is defined. In the "comparability" approach, the job of defining utility is left to an outside field, such as moral philosophy. The analytical work performed by the theory of social choice is limited to the problem of aggregation, that is, of trading-off gains and losses in well-being for different individuals. In the fairness approach, the construction of *what is being compared* is also justified *within* the theory, on the basis of axioms of fairness in particular. For instance, nothing a priori says that, in the distribution problem, the percentages of total resources that individuals would accept should play such a central role. It is only because of axioms such as Minimal Egalitarianism on Shares that they turn out to be relevant. In this approach, the set of axioms must, in one blow, define how to measure individual situations to be compared, and how to adjudicate conflicts of interests between individuals. In some cases it is actually possible to separate the role of different axioms into those which determine the measurement of individual situations and those which deal with inter-individual trade-offs. But the important point is that all this belongs to the same theory.

The best for the second-best

So far, I have clarified the similarities and differences between the more traditional "comparability" approach and the "fairness" approach. I shall now argue that the latter deserves to be applied. First, I will discuss the possibilities offered by this approach for the normative economic analysis of institutions, and argue that the design of optimal institutions, when it is based on fairness principles, should rely on this approach. Second, I will examine

[28] On interpersonal comparisons in general, see Fleurbaey and Hammond (2004).

the philosophical relevance of focusing on individual bundles and fairness conditions, and relate this approach to the fairness principles which appeared in recent philosophical theories of justice. This two-part analysis will not only establish a vague congruence between theories of justice and the ethical principles underlying the fairness approach, it will suggest that this approach provides a unique tool for the application of recent theories of justice to the design of just institutions.

The previous sections have focused on the definition of social preferences. The selection of institutions or policies on the basis of social preferences, roughly speaking, consists in finding the best option for the preorder formed by such social preferences in whatever set of feasible options is at hand. The celebrated "Arrow program" of social choice nicely distinguishes these two stages. In the first stage, each profile of preferences in the domain of the problem is aggregated into a social preference.[29] Then, once the actual profile and the actual feasible subset of alternatives are known, the social preferences corresponding to this profile are maximized on the feasible set in order to determine the social choice. In practice, the main difficulty in this optimization exercise is that the decision-maker often has an imperfect picture of the set of attainable options and of the profile of individual preferences.

The two issues (knowing the attainable set, knowing the profile) are linked, since the individuals may be expected to react to the rules of the game by pursuing their own interest rather than a collective goal, so that the attainable allocations depend on the profile of preferences. Whether this makes achievable states of affairs radically unjust, or simply puts a constraint on the design of just institutions, has been the topic of a recent controversy raised

[29] We have seen that the idea that feasibility considerations should be totally absent from this first stage may sometimes be contentious. But even when some information about feasibility is used in the definition of social preferences, there may still be some variation in the actual set of options that is considered in the final stage of decision, so that the general idea of this two-stage procedure remains useful.

by Cohen (2000). Here I am only interested in the search for *optimal* institutions in realistic conditions of human behavior, independently of whether such institutions are perfectly just.

Economists have defined several kinds of incentive-compatibility constraints, depending on the amount of information about the population profile that is available when writing the rules of the game. I will focus on the most relevant setting for the discussion of redistribution through devices such as taxes and benefits, social insurance and public assistance. This setting is usually called the "second-best" context. It consists in the following kind of assumption. The designers of redistributive mechanisms know the distribution of characteristics of the population, but do not know who has what characteristics in detail. They do not observe, at the individual level, personal characteristics such as educational background, physical strength, and other talents. Nor do they observe all of people's choices, such as their efforts at work or the time they spend working.

The second-best is opposed to the "first-best", which is the situation when the designers of the redistribution mechanisms know everything about everybody, so that they are only constrained in their control of the social state of affairs by the technical constraints of transformation and allocation of goods.[30] In the second-best context, redistribution is constrained, in addition, by the designers' lack of knowledge of individual characteristics, so that transfers have to depend only on observable variables such as earnings. Usual tax and subsidy instruments are of this imperfect kind. Nonetheless, the fact that the designers know the statistical distribution of all relevant data over the population is important, because it enables them to compare the consequences of different institutions. For instance, they can forecast that if the tax scheme is such and such, people with such and such characteristics will work and earn that much, yielding a global social state that can be evaluated if social preferences are well defined.

[30] This does not mean that this implies some authoritarian planning. As shown in famous theorems of welfare economics, the market may be used, supplemented by appropriate transfers.

This is where social preferences of the various kinds discussed above are helpful. Since the social state of affairs induced by any precise institution can be accurately forecast, social preferences on social states may be used to derive social preferences on institutions.[31] Once social preferences on institutions are obtained, it is a pure optimization exercise to find the best feasible institutions.

If one is interested in applying fairness principles to the design of institutions, the "fairness" approach appears as the natural tool to rely upon in order to elaborate a rigorous method for this purpose. This is because it defines social preferences that incorporate fairness principles and because such social preferences can then directly be incorporated in the optimization exercise described above. However, two objections, leading each to a different method, have to be discussed.

First, one may ask whether it is necessary to define social preferences over all allocations, as done in social choice in general. After all, one decision will be eventually made, leading to a precise social state of affairs (ignoring any intrinsic uncertainty). It would be enough to be able, in every possible context, to select the best allocations among those which are achievable through feasible institutions. This kind of objection has repeatedly been made against social choice in general. Notice, nonetheless, that selecting a subset of allocations still amounts to defining a complete preorder over the set of allocations, so that this is not formally an alternative approach.[32] What is really meant by this suggestion is that one could ignore fine-grained distinctions among bad allocations, such as required by the

[31] The astute reader will have noticed that this assumes that social preferences are anonymous. Since it is ignored who has what characteristics, social preferences may serve to evaluate the social state of affairs only if they focus on the distribution of individual situations, and do not want to favor Jones or Smith. As a counterexample, suppose that some particular redistributive scheme is favorable to hardworking people. If social preferences are not anonymous and want to favor Jones particularly, one cannot decide whether the redistributive scheme under consideration is good or not when one ignores whether Jones is hardworking or lazy.

[32] Selecting a subset of allocations is equivalent to defining the following social preferences: Any allocation that is selected is strictly preferred to any allocation that is not selected; all selected allocations are equivalent; all non-selected allocations are equivalent. For a detailed comparison of the "social preference" approach (that ranks all options) and the "social choice correspondence" approach (that selects a subset), especially with respect to the

Weak Pareto axiom. This idea of directly looking for the optimal second-best allocations is a priori a good one, but it seems hard to apply, for the following reason. Usually, fairness principles applied to allocations, as opposed to comparisons of allocations, are formulated in terms of some equality requested in some particular cases. For instance, consider the principle that "individuals with identical preferences should end up on the same indifference curve". The incentive-compatibility constraints that make up the second-best setting generally prevent such equalities to be achieved (for instance, the talented individual can decide to work less if taxed, thus retaining some advantage over the less talented). The usual trick on which one falls back in such a situation is to replace the equality requirement by a priority statement, which takes the form of… a comparison between allocations. For instance, a transfer principle may replace a full equality request, as in Transfer Principle for Equals, which says that some kind of equalizing transfer produces a social improvement. This is indeed comparing the post-transfer allocation to the pre-transfer allocation. Now, by falling back on such comparisons of allocations, one is essentially turning to the social choice approach proposed here.[33]

Another contender is the method that consists of defining a pure fairness ranking, ignoring Pareto considerations, and then, applying this fairness ranking to the subset of second-best efficient allocations only, in order to avoid ending up with an inefficient allocation. This method is not unreasonable, and can yield interesting results.[34] One may, however, wonder whether it is a radically different method, or a particular case of the method proposed here. Its main characteristic is that it entirely drops any Pareto requirement in the definition of the ranking. It would still have to discuss social preferences on the basis of axioms, and this is not fundamentally different from the "fairness" approach in general, apart

information they use about individual preferences, see two complementary viewpoints in Le Breton (1997) and Fleurbaey, Suzumura and Tadenuma (2005b).

[33] In the first-best context, this difficulty is absent, and this is how the impressive theory of first-best fair allocation (surveyed in Moulin and Thomson, 1997) has indeed been developed. Unfortunately, the many positive results of this theory are useless for second-best applications.

from this particular choice about Pareto, which should render the axiomatic analysis of social preferences more difficult rather than easier.[35]

It may also be emphasized that the full-fledged social preferences epitomized in the fairness approach can be useful under any kind of restriction about the feasible set, whereas the above two alternative approaches are much less powerful. Consider the "reform" setting, in which there is a status quo and only a handful of possible reforms. This is the actual setting in which policy decision-making is made in practice. With well defined social preferences, one just has to select the best reform (or to reject all of them if the status quo appears to be superior). The alternative approach which, as described above, directly selects the second-best option is silent there if, as it is unfortunately likely to be the case, none of the available options corresponds to the optimal second-best option. The other approach (the one involving a pure fairness ranking) would be applicable but would possibly yield bad decisions. It would select the option that is the most fair among those which are not Pareto-dominated by other options in this small set. But this might be a bad selection, if this option is very bad in terms of efficiency with respect to the whole set of conceivable options. With full-fledged social preferences that combine efficiency and fairness concerns, one can sometimes reject an option that is fair but very inefficient and prefer another one which is much more efficient even if slightly less fair.[36] This kind of trade-off is not possible with a social ranking that is exclusively focused on fairness.

To summarize, the main point of this section is that social preferences constructed in the "fairness" approach to social choice can readily be applied to the problem of choosing the

[34] An example of this approach is given by Nishimura (2003a,b), who applies a fairness ranking due to Chauduri (1986).
[35] I am not aware of any axiomatic result along this vein.
[36] Here is an imperfect but simple illustration of this, in the comparability case. In a two-individual population, the utility distribution (5,7) is more equal than (10,6), and is not Pareto-dominated by it, but it seems clear nonetheless that the latter is better. This is an imperfect illustration because in the comparability case there is an obvious remedy to this problem, which consists in examining distributions after sorting out utilities in increasing order (then one sees that (5,7) is Pareto-dominated by (6,10) and eliminated immediately). In a non-comparable

optimal institutions in second-best (and, as explained with respect to the reform setting, in third-best, fourth-best, etc.) contexts. As explained in the next section, this should be good news to philosophers who develop concepts of fairness. For lack of the fairness method, philosophers like John Rawls and Ronald Dworkin have resorted to expedients such as unspecified synthetic indices (e.g. Rawls' index of primary goods) or unspecified notions of distance to a first-best allocation (e.g. Dworkin's hypothetical market as a guide to tax institutions). The fairness approach provides theories of justice with a tool for applications, and in return such theories give additional relevance to the fairness approach.

Equality of resources from theory to tax

The initial motivation for developing concepts of fairness has been quite different in economics and in philosophy. In economics, the theory of fairness has been developed as an extension of general equilibrium theory and retained the same framework of multiple goods evaluated in a non-comparable way by individual ordinal preferences. While economists fret about interpersonal comparisons of well-being and feel more at ease with bundles and preferences mostly on positivist or behaviorist grounds, philosophers such as Rawls and Dworkin have come up with deeper ethical justifications for avoiding interpersonal comparisons of subjective wellbeing. They argue that social justice should be defined in terms of equality of resources provided to individuals, implying that individuals are left responsible for the way in which they transform resources into wellbeing (e.g., achievements and satisfaction). In the case when resources do not benefit equally to differently gifted individuals, personal talents may be counted among the “extended” resources made available to individuals, as in Dworkin's theory. The important point is that, for the allocation of resources, the only information about individual well-being that is relevant is their ordinal

setting, one can find more appropriate, albeit more complex, examples and the remedy is no longer available because sorting out utilities requires comparable figures.

non-comparable preferences, whereas individuals' responsibility for their ambitions and views of the good life makes it useless to consider any comparable index of satisfaction. The fact, for one individual, to have a lower satisfaction for any given amount of resources is then not a reason to give him more resources.

One might object that such responsibility should even make it superfluous to take account of ordinal preferences, and there is a sense in which this is correct, but there is another sense in which this would be absurd. It is correct that having particular ordinal preferences should not be a reason to receive more or less resources, overall, than others; but it would be absurd to let the precise distribution of particular goods disregard people's tastes, leading to gross inefficiencies. Personal responsibility for preferences should not be used in a way that undermines the Pareto principle. Individual preferences are therefore indispensable for the valuation of goods and the comparison of the value of bundles obtained by different individuals.

It is also important to dispose of another possible objection against the idea of disregarding comparable information about subjective satisfaction. It may indeed be confused with the claim that making people happy is not a valuable social goal, or even with the dreadful idea that preventing or relieving pain and suffering is not important. When it is said that “satisfaction” does not matter for social justice, it concerns the evaluation of one’s life, not the experiential content of one’s life. If pleasure and avoidance of pain are important dimensions of achievement for some people –they are for most people nowadays–, then such people will use their resources and opportunities to cater to these dimensions and the Paretian idea that their preferences should be respected will make this an important issue. The case of people who suffer from bad health or who are less good at transforming consumption into pleasure (or medical treatment into pain relief) can be addressed in terms of internal resources and handicaps which may indeed call for compensatory help. In contrast, a pure

dissatisfaction with one's life is not in itself a reason for obtaining a greater share of resources. Truly enough, dissatisfaction and suffering are often experienced together, but ideally they should be distinguished. Under equality of resources, therefore, someone who feels miserable because he has no more resources than the others (his goal in life being, say, to be a tycoon) will be eligible for help only because of the depression he may suffer as a result, not because of his lack of satisfaction as such.

Such theories of justice are very close, in their formulation, to the fairness idea that individuals should receive bundles of equal value, as assessed through their preferences. In this section I want to show the potentialities and advantages of a rigorous method starting from basic principles, deriving social preferences from the basic principles, and ending with the analysis of optimal institutions for such social preferences.

Dworkin is the author who makes the most use of fairness concepts such as the equal-split auction or the envy test.[37] But he is obviously at pains when he has to describe what the best redistributive scheme should be. His favorite theoretical reference is a hypothetical (veil of ignorance) insurance market on which individuals could buy insurance, including coverage against personal handicaps. But how does one relate the virtual outcomes of this hypothetical market to actual redistribution? The only answer that Dworkin proposes is that the redistribution system should try to come as close as possible to the hypothetical ideal. Now, how does one define closeness? This must involve a notion of distance, allowing a comparison of allocations. We are then not very far from the social choice problem.

But Dworkin's reference to the insurance market is somehow half-way between a general formulation of principles and a concrete description of optimal institutions. Dworkin also defines general principles. First, he relies on the envy test in order to check equality of resources at some basic level. He then goes on and adopts the equal-split auction as an ideal

[37] See Dworkin (1981, 2000).

mechanism which guarantees that the resulting allocation is envy-free. And he conceives the best application of the equal-split auction to the problem of personal talents as a formula of equal endowments on the hypothetical insurance market. In addition, he formulates two general principles in order to define equality of extended resources at a very general level. First, the *final* allocation of resources should be 'endowment-insensitive', in order to avoid any penalization of people with bad personal characteristics. Second, it should be 'ambition-sensitive', in order to reflect people's choices based on their particular preferences.

Now, relying on the concept of envy and the last two principles, it is possible to show that Dworkin's insurance mechanism is a very poor device. This has been shown by Roemer (1985), and I recently tried to analyze this problem in more detail.[38] In a nutshell, the conclusion of this analysis is that the hypothetical insurance mechanism is not endowment-insensitive because it implies transfers that may unduly penalize those whose personal talents and handicaps increase their productivity or decrease their marginal utility. It is also not ambition-sensitive in a satisfactory way since it makes the size of transfers depend on risk aversion (i.e. the purchase of insurance to cover the hypothetical risk of the hypothetical market) even in a context in which there is no intrinsic risk.

In two other papers, [39] François Maniquet and I have instead applied the two principles of endowment-insensitivity and ambition-sensitivity in a full-blown social choice analysis of the problem of redistribution when people have different labor-income preferences and unequal earning abilities. The results are interesting families of social preferences, which perform better than Dworkin's mechanism in terms of the two principles.

As an illustration, consider the following axioms, applied to this context where individuals are described by their labor-income preferences and their earning abilities. As in Dworkin's approach, it is assumed that earning ability is a characteristic which elicits social

[38] Fleurbaey (1998, 2002).
[39] Fleurbaey and Maniquet (2002, 2005).

help when it is low. The first axiom reflects the endowment-insensitivity requirement quite clearly, by requiring the avoidance of any inequality between individuals who differ only in their earning ability.

Compensatory Transfer: If two individuals have identical preferences and work the same amount, but have different consumptions, then it is socially at least as good to make a transfer that reduces consumption inequality between them (leaving labor quantities unchanged).

To prevent any equivocation, it must be emphasized that such an axiom is not about institutions but about a pure social choice comparison of allocations. The way in which such allocations might be obtained is irrelevant at this stage. This axiom invokes a hypothetical transfer, but at this stage one does not ask whether such a transfer is feasible or compatible with incentives. The question is only: If such a transfer *could* be done, *would* that be a good thing? The axiom answers that it would never be a bad thing.

It is this first axiom which Dworkin's mechanism miserably fails to obey, because the hypothetical insurance market will lead people to maximize their expected utility over all possible earning abilities that they may have, and since an expected utility is a weighted sum of utilities, this will lead to utilitarian-like, not egalitarian kinds of redistribution.[40] It will require, typically, putting the more talented on a lower indifference curve whenever that is possible.

The second axiom reflects the ambition-sensitivity requirement, by sanctioning the absence of redistribution when all agents have the same earning ability. The laisser-faire allocation is, then, such that unequal incomes only reflect individual preferences about leisure and income.

[40] For details, see Fleurbaey (2002).

Laisser-Faire in Equal-Skill Economies: If all individuals have the same earning ability, no redistribution is needed.

The third and final axiom is a mere separability[41] requirement whose purpose is to make sure that the social preferences will have a simple structure. It can also be justified on the grounds that individuals who are not affected by a change in the allocation of resources need not have a say in the evaluation of this change.

Separability: If an individual has the same labor-income bundle in two allocations, the social ranking of these two allocations should be the same if this individual did not exist.

These three axioms, combined with Weak Pareto and Weak IIA, entail that an allocation must be strictly preferred to another if its smallest "equivalent ability" is greater.[42] The "equivalent ability" measures, for any individual at any particular labor-income position, the earning ability that would provide the same satisfaction as her current position to this individual, in absence of any tax and transfer (that is, if this individual had to live with this earning ability in a laisser-faire economy). As an example, consider an individual who works and pays taxes. Suppose he was offered a job with a tax-free wage (and no restriction on his choice of labor time), and he would only be willing to accept it if he was offered an hourly wage of at least *w*. Then *w* measures his "equivalent ability". Now, the *smallest* equivalent ability of a given allocation is simply the smallest value of individual equivalent abilities, among the population, for this allocation.

[41] Roughly speaking, social preferences are separable when the evaluation of a change affecting a subgroup only does not depend on the particular (unchanged) situation of the rest of the population.
[42] See Theorem 1 in Fleurbaey and Maniquet (2002). This result is just an example. With a different set of axioms, and in particular by replacing Laisser-Faire in Equal-Skill Economies by other expressions of ambition-sensitivity, one derives slightly different social preferences, which can also claim to represent equality of resources. See Fleurbaey and Maniquet (2005) for details, and Fleurbaey (2002) for a non technical presentation of these alternative social preferences.

These social preferences, in other words, compare individual positions in terms of equivalent ability, and apply the maximin criterion[43] to such measures of individual situations. Let us compare this kind of result to Dworkin's approach. We have formulated axioms that encapsulate Dworkin's principles of endowment-insensitivity (Compensatory Transfer) and ambition-sensitivity (Laisser-Faire in Equal-Skill Economies). Dworkin's mechanism fails to satisfy the first axiom, and is known for not being endowment-insensitive. It does satisfy the second axiom, but in general makes the distribution of resources also depend on risk aversion, whereas in this simple context there is no intrinsic risk in the economy. In contrast, risk aversion plays no role in the definition of the "equivalent ability" social preferences. In summary, the fairness approach seems superior in terms of Dworkin's own principles. Additionally, it yields social preferences which can be used, as explained in the previous section, for a precise evaluation of institutions in the second-best context, whereas Dworkin's mechanism is only imperfectly comparable to real-life redistributive policies and therefore cannot yield similarly precise recommendations.

As an illustration of this point, let us briefly examine what kind of redistributive institutions would be optimal for such social preferences in the second-best context described in the previous section. Two facts are worth mentioning in particular.[44] First, if there are individuals with no productive ability, then the best redistribution scheme is the one which maximizes the minimum income guaranteed to all, and which may be distributed as a basic grant. But the second fact is that, if the minimal earning ability in the population (among actual, not equivalent, abilities) is greater than zero, then the best redistribution scheme no longer maximizes the basic grant. In this case, the individuals who have the minimal earning ability and work full time should not be taxed and should instead receive a subsidy, as in the

[43] More precisely, they apply the strict preference part of the maximin criterion. The formal results do not make precise what happens when two allocations have the same smallest equivalent ability. This has no importance in applications.
[44] See Fleurbaey and Maniquet (2002).

Earned Income Tax Credit system. More strikingly, they should actually be given the *greatest* subsidy of the whole population. In other words, the optimal tax schedule is then very favorable to the hardworking poor. More precise results about the contents of the tax and transfer policies can easily be derived when more information is available about the particular characteristics of a given population.

This comparison between Dworkin's approach and an alternative analysis based on similar ethical principles and the fairness approach to social choice lends support to the more general claim that, more than available alternatives, this methodology makes it possible 1) to derive precise social preferences; 2) in a rigorous (namely, axiomatic) way; 3) leading to precise institutional recommendations.

The "fairness" approach to social choice, insofar as it relies on non-comparable ordinal measures of individual well-being, is not appropriate for the application of all theories of justice. For instance, it cannot accommodate theories of equality of opportunities in which individuals are not responsible for their "utility functions". However, the "fairness" approach is not only suitable for theories of equality of resources. Any theory which declares that individual levels of satisfaction are irrelevant to issues of justice can turn to this approach for the study of economic applications. In addition, consider any theory that defines well-being in some more or less objective way, such as Sen's theory of functionings and capabilities.[45] It might be useful for such a theory to be able to proceed on the sole basis of an ordinal ranking of well-being for every individual. Usually the hardest part of such a theory is to make interpersonal comparisons. Whether the fairness concepts would translate into sound principles for such a theory remains to be seen and will not be explored here, but the general possibility of making social choice with a non-comparable notion of well-being is probably worth bearing in mind in many contexts.

Conclusion

Let me summarize the main points. First, there exists an approach to social choice which avoids resorting to interpersonally comparable measurements of well-being, and obtains possibility results by relaxing Arrow's axiom of IIA. This “fairness” approach does not only provide possibilities, but allows the analyst to impose substantial equity principles on social preferences.

Second, such social preferences can immediately be used in the selection of second-best institutions, under incentive-compatibility constraints. This is not important for the discussion of abstract ethical principles, but essential in the perspective of applications, and philosophers who have attempted to give institutional recommendations within their theories of justice have been clearly handicapped by the lack of a method.

The combination of these two points provides new tools for the application of theories of justice in which it is considered irrelevant to make the allocation of resources depend on the comparative levels of individual subjective satisfaction. An example of a rather substantial revision of Dworkin's theory with this approach has shown how one can proceed and obtain conclusions that are more rigorous, more precise, and even more faithful to the basic principles. This is but an example and by no means a definitive statement of an alternative theory. It is an illustration that calls for further explorations, in particular of more detailed and more complex settings of application. Such explorations require the joint expertise of specialists from philosophy, social choice and public economics, and this motivates the call made in the introduction for a closer collaboration between them. From such collaboration one may hope that justice and fairness concerns, because they will yield more precise recommendations, will gain more relevance in policy debates.

[45] This theory is usually described as requiring a unique index of capabilities that would apply uniformly to all

Acknowledgments. This paper is based on a part of a talk given at the conference "Procedural Approaches to Conflict Resolution", Bielefeld, Oct. 2001, and was presented at a workshop on equal opportunity, Bayreuth, Feb. 2002. I thank the participants at these two events, as well as participants at a philosophy seminar in the University of Montreal, for their reactions. I am grateful to François Maniquet for many discussions and a long collaboration on these issues, and to Peter Vallentyne and two referees for detailed comments and helpful suggestions on an earlier draft.

References

Arrow, K.J. 1951, *Social Choice and Individual Values*, New Haven: Yale University Press.

d'Aspremont, C. and L. Gevers 1977, "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies* 44: 199-209.

Chauduri, A. 1986, "Some Implications of an Intensity Measure of Envy," *Social Choice and Welfare* 3: 255-270.

Cohen G.C. 2000, *If You're an Egalitarian, How Come You're So Rich?*, Cambridge, Mass.: Harvard University Press.

Dworkin R. 1981, "What is Equality? Part 2: Equality of Resources," *Philosophy & Public Affairs* 10: 283-345

Dworkin R. 2000, *Sovereign Virtue. The Theory and Practice of Equality*, Cambridge, Mass.: Harvard University Press.

Fleurbaey M. 1998, "Equality among Responsible Individuals," in J.F. Laslier, M. Fleurbaey, N. Gravel and A. Trannoy (eds.), *Freedom in Economics*, London: Routledge.

Fleurbaey M. 2002, "Equality of Resources Revisited," *Ethics*, 113: 82-105.

individuals. The fairness approach suggests to consider the possibility of having personalized indexes reflecting each individual's preferences over her own functionings and capabilities.

Fleurbaey M. 2003, "On the Informational Basis of Social Choice," *Social Choice and Welfare* 21: 347-384.

Fleurbaey M. 2004, "Two Criteria for Social Decisions," Nuffield College Economics W.P. 2004-W27.

Fleurbaey M. 2005, "The Pazner-Schmeidler Social Ordering: A Defense," *Review of Economic Design* 9: 145-166.

Fleurbaey M. and P.J. Hammond 2004, "Interpersonally Comparable Utility," in S. Barbera, P.J. Hammond and C. Seidl (eds.), *Handbook of Utility Theory*, vol. 2, Dordrecht: Kluwer.

Fleurbaey M. and F. Maniquet 1996, "Utilitarianism versus Fairness in Welfare Economics," forthcoming in M. Salles and J.A. Weymark (eds.), *Justice, Political Liberalism and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge : Cambridge University Press.

Fleurbaey M. and F. Maniquet 2001, "Fair Social Orderings," forthcoming in *Economic Theory*.

Fleurbaey M. and F. Maniquet 2002, "Fair Income Tax," forthcoming in *Review of Economic Studies*.

Fleurbaey M. and F. Maniquet 2005, "Fair Social Orderings with Unequal Production Skills," *Social Choice and Welfare*, 24: 1-35.

Fleurbaey M., K. Suzumura and K. Tadenuma 2005a, "Arrovian aggregation in economic environments: How much should we know about indifference surfaces?," *Journal of Economic Theory* 124: 22-44.

Fleurbaey M., K. Suzumura and K. Tadenuma 2005b, "The informational basis of the theory of fair allocation," *Social Choice and Welfare* 24: 311-342.

Hammond P.J. 1987, "Social Choice: the Science of the Impossible?," in G.R. Feiwel (ed.) *Arrow and the Foundations of the Theory of Economic Policy*, New York: New York University Press.

Hansson B. 1973, "The Independence Condition in the Theory of Social Choice," *Theory and Decision* 4: 25-49.

K.-I. Inada 1964, "On the economic welfare function," *Econometrica* 32: 316-338.

Le Breton M. 1997, "Arrovian social choice on economic domains," in K. J. Arrow, A.K. Sen and K. Suzumura (eds.), *Social Choice Re-examined*, vol. 1, London: Macmillan and New-York: St. Martin's Press.

Moulin H. and W. Thomson 1997, "Axiomatic Analysis of Resource Allocation Problems," in K. J. Arrow, A.K. Sen and K. Suzumura (eds.), *Social Choice Re-examined*, vol. 1, London: Macmillan and New-York: St. Martin's Press.

Nishimura Y. 2003a, "Optimal commodity taxation for reduction of envy," *Social Choice and Welfare* 21: 501-527.

Nishimura Y. 2003b, "Optimal non-linear income taxation for reduction of envy," *Journal of Public Economics* 87: 363-386.

Pazner E. 1979, "Equity, Nonfeasible Alternatives and Social Choice: A Reconsideration of the Concept of Social Welfare," in J.J. Laffont (ed.), *Aggregation and Revelation of Preferences*, Amsterdam: North-Holland.

Pazner E. and D. Schmeidler 1978, "Egalitarian Equivalent Allocations: A New Concept of Economic Equity," *Quarterly Journal of Economics* 92: 671-687.

Plott C.R. 1976, "Axiomatic social choice theory: An overview," *American Journal of Political Science* 20: 511–596.

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

Risse M. 2001, “Arrow's Theorem, Indeterminacy, and Multiplicity Reconsidered,” *Ethics* 111: 706-734.

Roemer J.E. 1985, “Equality of Talent,” *Economics and Philosophy* 1: 151-187.

Sen A.K. 1979, “Personal Utilities and Public Judgements: or What is Wrong with Welfare Economics?,” *Economic Journal* 89: 537-58.

Sen A.K. 1999, “The Possibility of Social Choice,” *American Economic Review* 89: 349-378.

van Parijs P. 1995, *Real Freedom for All*, Oxford: Oxford University Press.

Young H.P. and A. Levenglick 1978, “A Consistent Extension of Condorcet's Election Principle,” *SIAM Journal of Applied Mathematics*, 35: 285-300.