Any non-welfarist method of policy assessment violates the Pareto principle: A comment*

Marc Fleurbaey†, Bertil Tungodden‡

September 2001

Discussion paper 23/01

1 Introduction

Suppose it is admitted that when all individuals prefer a state of affairs to another, the social preference should follow their unanimous assessment. Then it seems hard to avoid the conclusion that when all individuals are indifferent between the two states of affairs, so should the social judgment be one of indifference. Because if the social judgment was a strict preference in that case, it would remain so in the case when all individuals have a strict but slight preference in the other direction, unless the social preferences are discontinuous with respect to the state of affairs.

This is a simple idea, and it is part of the folk knowledge in welfare economics. Nevertheless, it has recently been formally stated in two papers, independently and simultaneously. But the two publications are very different in style and content.

In one case (Suzumura 2001), this simple fact is only a small part of a much more elaborate construction in which it is shown, under precise conditions, how all of the so-called Pareto principles[1] are equivalent. It is presented as a technical and precise formulation of an idea which has no profound philosophical or political implications, and it is published in a journal which is dedicated to the diffusion of this kind of results.

This is in stark contrast with the other publication of the same argument (Kaplow and Shavell 2001). Here, no further elaboration of this idea is at-

*This was written when the first author was visiting the Norwegian School of Economics. We should like to thank John Broome, Agnar Sandmo and Erik Øiolf Sørensen for useful comments. The usual disclaimer applies.

†University of Pau (France). Email: marc.fleurbaey@univ-pau.fr.

‡Norwegian School of Economics and Business Administration, Bergen (Norway). Email: bertil.tungodden@nhh.no.

[1]Namely, according to the terminology of the quoted paper: Weak Pareto (x is socially preferred to y if all individuals strictly prefer x), Pareto-Indifference (x is socially indifferent to y if all individuals are indifferent between x and y), Strong Pareto (x is socially preferred to y if all individuals weakly prefer x and at least one individual strictly prefers x), Full Pareto (the combination of Pareto-Indifference and Strong Pareto).

tempted, and the formal presentation is reduced to its minimum. But, although the formal result is very limited in content, the authors somehow manage to derive impressive political conclusions from it, arguing, in a nutshell, that it vindicates welfarism and excludes any consideration of fairness principles and any deviation from purely welfarist social welfare functions in political decision-making and institution design. And this appears in a generalist journal with a wide impact.

The problem, however, is that the conclusions of Kaplow and Shavell are gross overstatements, based on misrepresentations of such important notions as welfarism and fairness.

2 The result

The target of Kaplow and Shavell is what they consider *a common belief* among policy-makers, citizens and even some economists, namely "that it would be reasonable to include some extra-utility elements, of fairness or of justice, in the social welfare function (in addition to concerns about distribution, which welfarism admits)." (2001, p. 282) And they pretend to show "that, for any non-welfarist method of policy assessment (i.e. for any nonindividualistic social welfare function), there always exist circumstances in which the Pareto principle is violated." (*ibid.*) What they mean by the Pareto principle is the Weak Pareto condition, and what they mean by a nonindividualistic social welfare function is any social welfare function violating the Pareto-indifference condition.[2] Their formal result is, as described informally in the introduction above, that the Weak Pareto condition implies Pareto indifference (under some ancillary assumptions, notably a continuity condition).

On the basis of this result, their conclusions are, first, that social decisions should only rely on a traditional social welfare function of the following sort:

$$W(U_1(x), ..., U_n(x)),$$

where $U_i(x)$, for $i = 1, ..., n$, is individual i's utility in state x, and, second, that non-welfarist fairness principles should be discarded.

Their formal result is correct, but their conclusions cannot be derived from it unless one severely misinterprets the concepts of welfarism and of fairness.

3 What is welfarism?

Kaplow and Shavell seem to believe that the Pareto indifference condition captures the idea of welfarism. But that is not the case. Welfarism, as defined in welfare economics and social choice theory, is much more demanding, and requires that the social preference relation in the utility space is *the same for all possible profiles of individual utility functions.* For instance, the utilitarian brand of welfarism takes the sum of utility in all circumstances. And it is well

[2] See the previous footnote for definitions.

established that no version of the Pareto condition can entail welfarism by itself.[3] Hence, *nothing* in the paper of Kaplow and Shavell provides an argument in favour of welfarism understood in this standard way.

The Pareto-indifference condition means only that, *for a given* profile of utility functions $(U_1, ..., U_n)$, there exists a ranking of the space of utilities which represents transitive social preferences. But the Pareto-indifference condition does not at all force us to apply this ranking when considering a different profile of utility functions $(V_1, ..., V_n)$. For example, suppose that $U_i(x) = V_i(x)$ and $U_i(y) = V_i(y)$ for all i. A welfarist approach requires the ranking between x and y to be independent of the utility profile in question. But, of course, this does not follow from the Pareto-indifference condition, because this condition has nothing to say about the relationship between rankings based on different utility profiles. It is only a condition on how to rank alternatives for a given profile of utility functions, and similarly, for all the other Pareto conditions. Hence, these conditions do not at all exclude the possibility of taking into account *non-welfarist information* when moving from one utility profile to another.

What kind of non-welfarist information may one like to take into account when moving from one utility profile to another? A prominent answer is provided by a branch of welfare economics usually called the "theory of fair allocation" or the "theory of fairness". To this notion we now turn.

4 What is fairness?

As far as fairness is concerned, Kaplow and Shavell are shaping the discussion in a very peculiar way. They do not focus on the need for a theory of fairness in order to solve conflicts between people. Rather, they look at principles of fairness and justice that are valued in their own right, that is regardless of their consequences for individual welfare. And it is suggested that examples of such principles are that individuals' rewards should be based on merit, that punishment should fit the crime, and that horizontal equity and basic capabilities should be taken into account when making social evaluations. Their argument is that if a social evaluator accords weight to such ideas, then at least he or she would prefer a state of the world x to another y, if these principles are more respected in x than y and everyone is indifferent between x and y.

But this way of posing the problem misses the main point with theories of fairness, namely that in general they are valued *because they are considered appropriate ways of solving conflicts of interest between people.* The aim of most non-welfarist approaches is to challenge the distribution of *utility* as the framework for solving conflicts. Hence, these principles are usually not suggested as elements of fairness that should be taken into account *in addition to* what Kaplow and Shavell name distributive considerations, that is, considerations based on utility, but rather that they should be used *instead of* such considerations.

[3] See d'Aspremont and Gevers (1977) and, for a survey, Sen (1986).

An interesting question, then, is whether one can use non-welfarist concepts of fairness in order to solve cases of conflict and at the same time satisfy the Pareto principles. As we have alluded to above, the "theory of fair allocation" provides a positive answer to this question.[4] More specifically, this theory says that one should pay attention to the change in the *overall* preference structure of individuals when considering the possibility of revising the ranking of alternatives when there is a change in the utility profile. Essentially all allocation rules studied within this branch of welfare economics are required to select only Pareto-efficient allocations, and most of these allocation rules do not make any distinction between Pareto-indifferent allocations. But they are not based on the maximization of classical social welfare functions, and, actually, most of them just ignore utilities and rely only on individual preferences. As an example, consider social preferences favoring, as the optimal social state, the egalitarian Walrasian equilibrium where all agents have the same budget. Such preferences satisfy the Pareto-indifference condition in the sense that if an allocation is feasible and is Pareto-indifferent to an egalitarian Walrasian equilibrium, it is also an egalitarian equilibrium and is therefore selected as well. And it has recently been shown that this allocation rule also maximizes *non-welfarist* social preferences satisfying all the Pareto principles.[5] But, as is well known, the egalitarian equilibrium cannot be rationalized as the optimal state for a classical social welfare function,[6] because the welfarist approach does not allow us to include information on the overall preference structure of individuals when ranking alternatives.

This is not to say that the different Pareto principles do not constrain the possibility to use non-welfarist reasoning when solving cases of conflict. They do, as shown by Sen (1970) and elaborated on by, among others, Gibbard (1979), Moulin and Thomson (1997), Brun and Tungodden (1999). But these considerations cannot be captured in any interesting way by the framework of Kaplow and Shavell. Hence, it is a major overstatement when they claim that "[o]ur conclusion, by contrast [to the impossibility result of Sen(1970)], is general in that it applies to any form of non-welfarism" (Kaplow and Shavell 2001, p. 282). Their result only applies to a small subset of non-welfarist theories, and for these theories it deals with a trivial tension between the Pareto principles and a requirement of continuity.

5 What is the Pareto principle?

There is a third misleading message in Kaplow and Shavell (2001). They expose Sen's (1985) theory as an example of a theory which, by introducing functionings and capabilities, is going away from welfarism and the Pareto principles.[7]

[4] For a recent survey, see Moulin and Thomson (1997).

[5] See Fleurbaey and Maniquet (1996). In these non-welfarist social preferences, the convex hull of the union of the individual upper contour sets plays a central role. The computation of these social preferences does not make use of utilities, and relies only on individual preferences.

[6] Except in very special cases. See Eisenberg (1961).

[7] Rawls' theory (1971), which focuses on primary goods, is a similar example.

By giving this example they apparently mean that their argument somehow questions the soundness of such a theory. That Sen's theory is non-welfarist is correct. That it should be shaken by Kaplow's and Shavell's argument is another story. Although this is not very explicit, they apparently believe that the Pareto principle should without question be conceived in terms of the satisfaction of individuals' ordinary preferences. But most philosophers and many economists have questioned the soundness of individual ordinary preferences as a basis of social evaluation. Ordinary preferences are usually misinformed, unreflected, myopic and sometimes antisocial. Many arguments of that sort lead to the conclusion that it is not obvious how to conceive the most appropriate notion of individual well-being.

Sen's theory, and its likes,[8] is constructed precisely on the basis of a thorough evaluation of what should be the proper *focus of social decisions*, as far as the definition of individual welfare is concerned. If the conclusion of this evaluation is that individual welfare should be measured in terms of capabilities, then there is not much reason to worry about violations of the Pareto principle with respect to ordinary individual preferences. In the computation of capabilities, proper weight is supposedly given to the satisfaction of individual preferences, among other kinds of functionings.

A theory which redefines individual welfare in some other way than in terms of utilities can, and probably should, redefine the Pareto principle similarly. Social preferences, according to such a theory, should indeed favor an increase in individual welfare as it is defined within this theory. And the advocates of such non-welfarist theories have been careful, of course, to adopt this revised Pareto principle.[9]

In other words, the point of these non-welfarist theories is usually less to go against the Pareto principle than to look for the appropriate definition of individual welfare to which the Pareto principle itself should be applied. It is then quite awkward and naive to criticize non-welfarist theories on the basis of the ordinary Pareto principle.

6 Conclusion

Suppose that someone told you that we should assign importance to the preservation of the Grand Canyon beyond its impact on people's welfare, and that there should be cases where we accept a trade-off between the intrinsic value of Grand Canyon and the welfare of people. Would it come as a surprise that we would then have to go against the welfare of people on some occasions?

This is basically the story told by Kaplow and Shavell, and which they use to vindicate welfarism and castigate fairness. Although their formal result is true (trivially so), there are three main mistakes in their bold conclusions. First, it is not true that the Pareto principles (even applied to ordinary preferences) forbid any departure from classical social welfare functions. Second, it is not true that

[8] Such as Broome (1991).

[9] Broome (1991, p. 155) explicitly names this revised version the principle of personal good.

concepts of fairness are generally incompatible with Pareto principles (and that they deal mostly with non-conflictual situations). Third, it is not true that the Pareto principle should unquestionably be conceived in terms of the satisfaction of ordinary preferences. These three mistakes rely on a misrepresentation of the notions of welfarism, fairness, and of the basic idea of the Pareto principles.

The developments of welfare economics and social choice theory in the last decades, notably in the theory of fairness and non-welfarist theories of justice, have not solved all problems and can rightly be criticized on many grounds. But they deserve to be studied in depth, and cannot be dismissed as easily as suggested by Kaplow and Shavell.

References

[1] Broome J. 1991, *Weighing Goods*, Oxford University Press.

[2] Brun B. C., B. Tungodden 1999, 'Non-welfarist theories of justice: Is "the intersection approach" a solution to the indexing impasse?', forthcoming in *Social Choice and Welfare*.

[3] d'Aspremont C., L. Gevers 1977, 'Equity and the informational basis of collective choice', *Review of Economic Studies* 44: 199-209.

[4] Eisenberg E. 1961, 'Aggregation of utility functions', *Management Science* 7(4): 337-350.

[5] Fleurbaey M., 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 University Press.

[6] Gibbard, A. 1979, 'Disparate Goods and Rawls' Difference Principle: A Social Choice Theoretic Treatment', *Theory and Decision* 11: 267-288.

[7] Kaplow, L. and S. Shavell 2000, 'Notions of Fairness versus the Pareto Principle: On the Role of Logical Consistency', forthcoming *Yale Law Journal.*

[8] Kaplow L., S. Shavell 2001, 'Any Non-welfarist Method of Policy Assignment Violates the Pareto Principle', *Journal of Political Economy* 109 (2): 281-286.

[9] Moulin H., 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.

[10] Rawls, J. 1971 *A Theory of Justice,* Harvard University Press.

[11] Sen, A. K. 1970 'The Impossibility of a Paretian Liberal', *Journal of Political Economy* 78: 152-157.

[12] Sen A. K. 1985, *Commodities and Capabilities*, Amsterdam: North-Holland.

[13] Sen A. K. 1986, 'Social choice', in K. J. Arrow and M. D. Intriligator (eds), *Handbook of Mathematical Economics*, vol. 3, Amsterdam: North-Holland.

[14] Suzumura K. 2001, 'Pareto principles from Inch to Ell', *Economics Letters* 70: 95-98.