Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle: Reply

Louis Kaplow and Steven Shavell
Harvard University and National Bureau of Economic Research

In Kaplow and Shavell (2001), we show that any non-welfarist method of policy assessment violates the Pareto principle. In particular, we demonstrate that if a social welfare function from complete descriptions of the world to the real numbers does not depend *solely* on individuals' utilities, and also satisfies a continuity condition regarding a single good, then there must exist situations in which the function rates one state of the world higher than another even though all individuals strictly prefer the second state of the world to the first.

As stated in our article, the import of this conclusion concerns the general use of non-welfarist principles by the public at large, policy makers, and philosophers—for instance, the notion that rewards should be based on merit or that the punishment should fit the crime—and the occasional use of non-welfarist principles by economists—such as the view that horizontal equity should influence tax policy or that individuals' capabilities rather than their utilities should be the focus of social decisions. Our result implies that none of these non-welfarist approaches can be accepted if one endorses the Pareto principle.

The thrust of Fleurbaey, Tungodden, and Chang's (2003) comment is that our result, although valid, relies on what they regard as strong assumptions and thus does not apply to every non-welfarist principle. They acknowledge that our result applies "if we consider only social choices based on a unique profile of individual utility functions in society" (p. 1383), and they go on to say that "welfarism, however, as this concept is understood in social choice theory, requires that the social preference relation in utility space (represented by the function W) be the same for all possible profiles of individual utility functions."

In other words, Fleurbaey et al.'s main point is that many social choice theorists define welfarism as embodying not just one but rather two types of restrictions: (1) that social policy decisions in a given society (which has only a single profile of individual utility functions) must depend only on individuals' utilities, and also (2) that even if social

choices in every society are purely welfarist as in restriction 1, the method of social choice may not vary across societies in a manner that depends on non-welfare information. Our result was aimed at restriction 1, and Fleurbaey et al. have no quarrel with us in this respect. Their objection is that our result can be understood to reach restriction 2 as well—thereby additionally ruling out the possibility of different social welfare functions for different societies—only if one is willing to make assumptions such as transitivity and independence (p. 1383).

Our response to Fleurbaey et al.'s argument is that we agree, but it is irrelevant given the purpose of our article.[1] The reason is that restriction 1 is patently violated by all the non-welfarist views of concern to us: the whole range of notions of fairness that are commonly employed by policy makers, the public, and philosophers from Aristotle to Kant to the moderns. That is, the principles in question all insist that a given society's decisions should not depend solely on information regarding individuals' utilities. We explained this in our article (p. 283, with formal elaboration in n. 5) with reference to the principle that the punishment must fit the crime, which clearly is not indifferent between two regimes with identical utilities but with punishments always fitting the crime in one regime and never doing so in the other.[2] That the vast multitude of non-welfarist principles at issue all have this property—of assessing alternatives for a given society based at least in part on non-utility information—not only is obvious but is also made explicit in our article, is elaborated at length in our related book (Kaplow and Shavell 2002) that we cite, and is not disputed by Fleurbaey et al. Put another way, none of the non-welfarist principles that are often offered in criticism of standard economic policy analysis would be satisfied if every society were, internally, purely welfarist with respect to literally every social decision, as long as the differences among societies' social welfare functions reflected non-welfare information in some particular fashion.

Finally, we observe that Fleurbaey et al.'s suggestion that our result is not original is misleading; in fact, the very theorists they cite in this

[1] Although not our focus, we do wish to register our belief that the assumptions of ours that Fleurbaey et al. highlight in challenging the broader, cross-society reach for our result are, upon reflection, compelling ones, which would be accepted by most philosophers who advance non-welfarist views. We elaborate on the reasons for this stance in Kaplow and Shavell (2000) in response to related prior arguments of Howard Chang, and we note that Fleurbaey et al. do not address these considerations.

[2] Indeed, this explanation and formal elaboration immediately follow the very observation of ours that Fleurbaey et al. criticize as not accurately embodying welfarism (p. 1383). This observation indicates that, under any non-welfarist principle, there must exist a case in which two regimes will not be assessed identically even though every individual's utility is identical in the two regimes.

regard hold a different view regarding the significance of our contribution.[3]

References

Campbell, Donald E. 2002. "On the Possibility of Efficient but Non-welfarist Policy Analysis." Manuscript. Williamsburg, Va.: Coll. William and Mary.

Campbell, Donald E., and Jerry S. Kelly. 2002. "Impossibility Theorems in the Arrovian Framework." In *Handbook of Social Choice and Welfare*, vol. 1, edited by Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura. Amsterdam: Elsevier Sci.

Fleurbaey, Marc, Bertil Tungodden, and Howard F. Chang. 2003. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle: A Comment." *J.P.E.* 111 (December): 1382–85.

Kaplow, Louis, and Steven Shavell. 2000. "Notions of Fairness versus the Pareto Principle: On the Role of Logical Consistency." *Yale Law J.* 110 (November): 237–49.

———. 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle." *J.P.E.* 109 (April): 281–86.

———. 2002. *Fairness versus Welfare.* Cambridge, Mass.: Harvard Univ. Press.

Suzumura, Kotaro. 2001. "Pareto Principles from Inch to Ell." *Econ. Letters* 70 (January): 95–98.

———. In press. "Welfarism, Individual Rights, and Procedural Fairness." In *Handbook of Social Choice and Welfare*, vol. 2, edited by Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura. Amsterdam: Elsevier Sci.

[3] Fleurbaey et al. assert that "others have derived similar results" (p. 1382), citing Suzumura (2001) and Campbell (2002). However, Suzumura's paper was written with knowledge of, and his results were derived from, an earlier version of our paper (see Suzumura, in press). In addition, Suzumura's *Handbook of Social Choice and Welfare* survey refers to our article as "important" and as posing a "devastating question" regarding all non-welfarist values. Campbell (2002) is a postdated working paper that makes express reference to our article, beginning in its abstract. Furthermore, Campbell and Kelly's (2002) own handbook survey describes our article as offering a "striking new impossibility theorem" (p. 80).