# WHAT (IF ANYTHING) CAN ECONOMICS SAY ABOUT EQUITY?

*Daniel A. Farber\**

Does economics have anything to teach us about the meaning of fairness? The leading practitioners of law and economics disagree. Judge Richard Posner argues that economics is largely irrelevant to distributive issues.[1] Posner maintains that the most useful economic measure of social welfare is cost-benefit analysis (which he calls wealth maximization). But, he observes, this economic measure "ratifies and perfects an essentially arbitrary distribution of wealth."[2] Given an ethically acceptable initial assignment of wealth, rules based on economic efficiency may have some claim to be considered fair.[3] On the critical issue of distributional equity, however, Posner apparently believes that economics has little to say.

In contrast, Professors Louis Kaplow[4] and Steven Shavell[5] believe that economics can teach us nearly everything about equity. In *Fairness Versus Welfare*, they argue that there is only one viable notion of equity: resources should be distributed so as to maximize overall social welfare. As we will see, the full import of this argument is ambiguous. On the one hand, Kaplow and Shavell expressly

---

1. Richard A. Posner, *Norms and Values in the Economic Approach to Law, in* NORMS AND VALUES IN LAW AND ECONOMICS (Aristides Hatzis ed., forthcoming 2003) (article manuscript on file with author).

2. *Id.* (article manuscript at 18).

3. *Id.* (article manuscript at 17) (asserting that almost everyone would agree *ex ante* to common law rules based on wealth maximization). For a summary of the objections to Posner's views, see BRIAN BIX, JURISPRUDENCE 125-26, 189-91 (2nd ed. 1999). I have previously argued that Posner's claim holds only under certain circumstances, namely, when legal rules do not strongly affect preferences or the distribution of wealth. *See* Daniel A. Farber, *Economic Efficiency and the Ex Ante Perspective, in* THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 54 (Jody S. Kraus & Steven D. Walt eds., 2000) [hereinafter Farber, *Ex Ante Perspective*].

4. Professor of Law, Harvard University.

5. Samuel R. Rosenthal Professor of Law and Economics, Harvard University.

concede that multiple ways of calculating social welfare might exist, so we might have to look beyond economics to determine the right one. On the other hand, much of their argument is implicitly predicated on a specific social welfare function, and in a footnote they give the argument for adopting this function universally.[6] If a unique social welfare function is given, economic analysis would completely resolve all equity issues under their approach.

Thus, although they do not say so explicitly, the book can be read to endorse a single definition of equity based on economic analysis. This reading would completely eliminate any independent role for judgments about equity. At the least, however, Kaplow and Shavell believe that economics can restrict value judgments about equity to a single, sharply defined place in policy analysis: the choice of an appropriate social welfare function. Posner, in turn, sees only modest merit in the Kaplow and Shavell theory of social welfare, except to the extent that it reduces in practice to his own theory of wealth maximization.[7].

Between Posner on one side, and Kaplow and Shavell on the other, there is a fundamental difference in visions of the normative side of law and economics. In one view, questions of equity fall outside of the domain of economic analysis; in the other, they may be entirely subsumed by it. Neither view is entirely satisfactory. Like Posner, I doubt that questions of equity can be reduced to economics, but like Kaplow and Shavell, I do think that economics has something useful to contribute to our understanding of equity.

In this Review, I do not directly defend other definitions of equity against Kaplow and Shavell's criticisms. Rather, I criticize their conception of social welfare, which is the basis of their theory.[8] My conclusion is that their version of welfarism is not a viable alternative. Thus, we are not in a position to reduce all ethical questions (other than choice of a social welfare function) to welfare economics. At least for those who believe that "it takes a theory to beat a theory," fairness norms remain unvanquished.

---

6. *See infra* text accompanying notes 20-22.

7. *See* Posner, *supra* note 1 (article manuscript at 19-20). As Posner's book blurb for *Fairness versus Welfare* shows, he also seems pleased at the challenge the book offers to moral philosophers, his own bête noir.

8. For other critiques of their theories, see Howard F. Chang, *A Liberal Theory of Social Welfare: Fairness, Utility, and the Pareto Principle*, 110 YALE L.J. 173 (2000); Ward Farnsworth, *The Taste for Fairness*, 102 COLUM. L. REV. 1992-93 (2002); David A. Hoffman & Michael P. O'Shea, *Can Law and Economics Be Both Practical and Principled?*, 53 ALA. L. REV. 335 (2002); LEWIS A. KORNHAUSER, A WEAVED UP FOLLY? PREFERENCE, WELL-BEING AND MORALITY IN SOCIAL DECISIONS (N.Y.U. Ctr. for Law & Bus., Working Paper No. 01-009, 2001), *available at* http://papers.ssrn.com/abstract=286772; and Joseph William Singer, *Something Important in Humanity*, 37 HARV. C.R.-C.L. L. REV. 103 (2002).

Briefly, the welfarist program advocated by Kaplow and Shavell has three fundamental flaws. First, except in the easiest cases where a decision benefits everyone, welfarism as such does not determine the outcome. Rather, the outcome depends *entirely* on the specific choice of a social welfare function. Thus, we need some standards for choosing the function. But these standards cannot be selected on welfarist grounds, because we have not yet chosen a social welfare function. As a general matter, non-welfarist fairness norms are needed to select the social welfare function ("SWF"), and this selection in turn will drive the ultimate social decision.[9]

Second, some set of procedures will be needed for society to decide on distributional norms. These procedures, which amount to a constitutional framework, must themselves be chosen on nonwelfarist grounds. Thus, the core of constitutional law is immune to welfarism. The only way to avoid this argument is to assume that only one valid SWF exists, so that no procedure for choice is necessary. Kaplow and Shavell implicitly resolve this dilemma by assuming that the SWF will be utilitarian — that is, that it merely adds individual utilities.

But choosing a utilitarian SWF exacerbates the third problem with welfarism, which is the difficulty of defining and measuring utility. Utilitarianism requires a high degree of interpersonal comparability of utility. So far, no one has successfully provided a definition of utility that is morally compelling (so that it is something that society has good reason to maximize), let alone a definition that provides a practical method for assigning quantitative values to utility levels. At the very least, we are a long way from being able to use utilitarianism as a practical method for resolving specific disputes about equity. Thus, Kaplow and Shavell's version of welfarism does not provide a satisfactory methodology for analyzing equity. It is at best premature — and more likely impossible — to try to reduce all issues of equity to welfare economics, as they seek to do..

Nevertheless, I will also suggest that Posner somewhat underestimates the possible contributions of economics to resolving fairness issues in law. For example, in the field of tax law, Kaplow and

---

9. A similar point is made in a recent article, Michael B. Dorff, *Why Welfare Depends on Fairness: A Reply to Kaplow and Shavell*, 75 S. CAL. L. REV. 847 (2002). But Dorff's treatment is limited to unusual types of social welfare functions, which may not even fit within Kaplow and Shavell's definition. These functions are not strictly increasing in individual utility, so that some improvements in individual utility are not counted at all in determining social welfare. *Id.* at 879-84. Kaplow and Shavell would reject Dorff's anarchistic and deontological functions, *id.* at 883, because these functions cannot be calculated purely on the basis of utility data about individuals, but also require data about the causes of utility changes. (Kaplow and Shavell do allow the use of non-utility information about individuals, but only to choose the Social Welfare Function ("SWF"), not to calculate the value of that function once it has been chosen. *See* p. 24 n.15.) The anarchistic function also violates the Pareto standard. The argument made in the present Review is much more general. In addition, Dorff does not consider the implications of his argument for constitutional rules.

Shavell's own previous work is part of an impressive body of scholarship elucidating the role of equity in taxation. Moreover, if we are willing to lower our sights a little, and ask about equity in disputes involving small numbers of individuals rather than about overall societal fairness, economics may have more to contribute. Economists and game theorists have devised some intriguing mechanisms for resolving distributive disputes. Economics cannot tell us precisely how to define equity, but it can help us to design mechanisms for achieving our notions of fairness, as well as sometimes helping us refine those concepts. Economic methodology is about how to go about maximizing some relevant attribute, but the attribute in question does not have to be efficiency. If what we want is to maximize equity, or some combination of equity and efficiency, economics can help us do that as well.

One note about terminology before we begin. Kaplow and Shavell use "fairness" as a term of art, to mean any value judgment that is *not* based on utility information. To avoid confusion, I will use fairness in their somewhat unusual sense as excluding utility-based standards. "Equity" will be used in a more encompassing sense to include any standard for determining the just distribution of resources, whether or not based purely on utility. Thus, for purposes of this Review, equity includes determinations about the overall distribution of utility, as well as including fairness in the Kaplow-and-Shavell sense of nonutility based equity.

Part I summarizes Kaplow and Shavell's argument against fairness. Part II turns to the problem of defining a social welfare function, while Part III probes the concept of individual utility. Finally, Part IV considers some uses of economic analysis and game theory in identifying methods for achieving equity.

## I.   WELFARISM, ECONOMIC EFFICIENCY AND EQUITY

Before considering the approach favored by Kaplow and Shavell, we should review the more conventional approaches taken by practitioners of law and economics. Posner's view of the relationship between equity and economic analysis is undoubtedly the conventional one among practitioners of law and economics.

The starting point for analysis is the idea of individual utility. Conventionally, "utility" is used as an arbitrary index for what is really an ordinal ranking of individual preferences; all that matters is the relative ranking rather than the actual magnitudes, which merely serve as a mathematical convenience. A more refined determination of utility (von Neumann-Morgenstern utility) can be calculated on the basis of individual choices between uncertain alternatives, provided certain assumptions are satisfied. Von Neumann-Morgenstern utility levels come somewhat closer to being meaningful quantitative

measurements than other utility functions. Because only the ratios between an individual's utilities are fixed by this procedure rather than their absolute magnitudes, however, we cannot use the von Neumann-Morgenstern procedure as a basis for comparing the utility levels of two different individuals.[10]

Since utility comparisons between individuals are not meaningful under the conventional approach, we can compare different states of the world only in a qualitative way. Economists most often use the Pareto standard to define efficiency.[11] Consider two states of the world, *A* and *B*. We say that *A* is Pareto-superior to *B* if a motion to shift from *B* to *A* would pass without dissent — that is, *A* ranks higher than *B* in the preference ranking of at least one person, who votes for the change, and the others are at worst indifferent to the change, so no one else objects. Pareto superiority is an intuitively appealing standard, since at least one person is better off because of the change and no one else is hurt. Essentially, the Pareto standard avoids the need for interpersonal comparisons by giving each individual a veto over changes. Because it seems to avoid any controversial value judgments, economists have favored the Pareto standard as the basis of what has been called the "new welfare economics."[12]

In reality, however, Pareto improvements are often hard to find. Because unanimous consent is unlikely to exist for changes in legal rules, law and economics practitioners often fall back to a broader but less compelling standard, Kaldor-Hicks efficiency. Alternative *A* is Kaldor-Hicks superior to alternative *B* if some people are better off with *A* and would be willing to fully compensate the "losers," regardless of whether or not such compensation is ever actually paid (pp. 458-59). The Kaldor-Hicks standard has two drawbacks, one technical and the other moral. The technical drawback is that the standard can be indeterminate — if sufficiently large wealth shifts are involved, *A* can be more efficient than *B* while *B* is also more efficient than *A*.[13] The moral objection is that Kaldor-Hicks ignores equity entirely, so that some people may be much worse off even though a change is economically efficient. Kaldor-Hicks improvements are arguably fair when everyone had an equal chance beforehand of being

---

10. The use of von Neumann-Morgenstern utilities allows us to determine the utility of an uncertain outcome, which is merely the probability of the outcome times its utility if it does occur. For an introduction to utility functions, see JAMES HENDERSON & RICHARD QUANDT, MICROECONOMIC THEORY: A MATHEMATICAL APPROACH 8-18, 56-61 (3d ed. 1980).

11. *See* MARK SEIDENFELD, MICROECONOMIC PREDICATES TO LAW AND ECONOMICS 49 (1996).

12. *See* DAVID A. STARRETT, FOUNDATIONS OF PUBLIC ECONOMICS 10 (1988).

13. *See* ANDREU MAS-COLELL ET AL., MICROECONOMIC THEORY 831 (1999).

on the winning side, but this is not always a realistic assumption.[14] When the Pareto principle does not apply and Kaldor-Hicks seems inadequate to deal with the issue of fairness, the conventional view requires us to look outside economics for guidance.

This brings us to the welfarist approach favored by Kaplow and Shavell, which is a variant of what has been called the "old" welfare economics.[15] As they explain, their approach also begins with the concept of individual utility. The utility function expresses the individual's ordering of outcomes, with expected utility being used in cases of uncertainty (pp. 24-25 n.15). Thus, the utility index "incorporates in a positive way everything that an individual might value — goods and services that the individual can consume, social and environmental amenities, personally held notions of fulfillment, sympathetic feelings for others, and so forth" (p. 18). Individuals may be mistaken about how outcomes will affect them. Hence, the utility index should relate to "what they would prefer if they correctly understood how they would be affected — rather than to individuals' well-being as reflected in their mistaken preferences" (p. 23).

Kaplow and Shavell do not limit themselves to the Pareto and Kaldor-Hicks standards in comparing different states of society. Instead, they base such comparisons on a social welfare function, which they define formally as a monotonic function of individual utilities. (This simply means that, all else being equal, an increase in any individual's utility increases the level of social welfare.) Once the SWF has been set, we need merely pick the alternative with the higher value. The key point is that *only* individual utilities count in determining social welfare (rather than, for instance, individual rights). Thus, any two outcomes that produce the same levels of utility are considered identical, whatever the sources of the utility may be. But in other respects, Kaplow and Shavell maintain, their approach is very accommodating, because there are many possible choices of SWFs. As they emphasize, a SWF can be "*any* increasing function of individuals' utilities."[16] One possibility is classic utilitarianism (where we add individual utilities), but we might also choose a SWF that gives more weight to the utility of the worse-off members of society (p. 27). Defining a SWF requires that we be able to compare individual utility levels, but Kaplow and Shavell are unfazed by the controversy over whether this is possible (p. 27 n.20).

14. Pareto optimality is defined in HENDERSON & QUANDT, *supra* note 10, at 286-91. Under the rubric of "wealth maximization," the Kaldor-Hicks standard is discussed in Posner, *supra* note 1 (article manuscript at 13-17). Wealth maximization is the same as Kaldor-Hicks, except that changes in utility are measured by their monetary equivalents. *See* SEIDENFELD, *supra* note 11, at 55.

15. *See* STARRETT, *supra* note 12, at 9, 11.

16. P. 24 n.15. As I discuss in Part II, however, much of their discussion effectively presumes that the SWF must be utilitarian.

Kaplow and Shavell also assume that it is irrelevant which utility levels are associated with which specific individuals (pp. 25 n.16, 26 n.18, 27 n.20). This assumption is much less innocuous than it may seem. In effect, this assumption rules out the possibilities that we might view some people as having morally superior preferences than others, or that we might consider some people as being more or less deserving than others because of their character or past conduct. Mother Theresa's preferences would count just as much in the SWF as Hitler's, no more and no less.

Welfarism may strike some people as insufficiently sensitive to issues of equity. If it can be implemented, however, it does offer a clear improvement over the Kaldor-Hicks standard because it takes some account of distributional issues. Under Kaldor-Hicks, essentially all that matters is maximizing social wealth. Holding constant the combined amount of wealth held by society as a whole, we have to count as equal (a) the state of the world where wealth is evenly distributed and (b) the state of the world where everything is owned by Bill Gates. In contrast, welfarism provides a basis for judging between different distributions of wealth. With plausible choices of the utility functions and the SWF, the assignment of wealth to a single individual will be rejected in favor of equal distribution.[17]

Still, there are aspects of equity that seem to be lost by the welfarist approach, such as the intrinsic value of human rights. Kaplow and Shavell maintain that the welfarist properly excludes these values. The SWF itself can incorporate certain elements of equity — those relating to the distribution of individual utilities as opposed to their sum — but Kaplow and Shavell argue against any additional role for moral judgments once the SWF has been chosen. In particular, they reject any notion of fairness that gives "weight to factors that are independent of individuals' well-being" (p. 44; emphasis omitted). For example, welfarism precludes libertarian views, in which some societal improvements might be blocked because they violate individual rights (p. 26 n.18).

The core of the argument is that use of any intrinsic moral factors would conflict with the Pareto standard. Kaplow and Shavell argue that "individuals will be made worse off overall whenever consideration of fairness leads to the choice of a regime different from that which would be adopted under welfare economics because, by definition, the two approaches conflict when a regime with greater overall well-being is rejected on grounds of fairness" (p. 52). Thus,

---

17. For example, if the marginal utility of money decreases and the social welfare function is additive — the classic utilitarian view — then a more equal distribution of wealth will always be favored. Thus, compared with the conventional economic approach, welfarist analysis is a great improvement in its handling of questions of equity.

anyone who believes in intrinsic moral values must be prepared to sacrifice human welfare.

The conflict is particularly clear when all individuals are identically situated. In such "symmetrical" situations, it is *"always* the case that everyone will be worse off when a notion of fairness leads to the choice of a different legal rule from that chosen under welfare economics" (p. 52). The explanation is simple:

> Because everyone is identically situated, whenever welfare economics leads to the choice of one rule over another, it must be that everyone is better off under the preferred rule. Hence, whenever a notion of fairness leads one to choose a different rule from that favored under welfare economics, everyone is necessarily worse off as a result. (p. 52)

Thus, when everyone is situated identically with respect to a rule, any independent consideration of fairness such as individual rights necessarily violates the Pareto principle.

In an important earlier article,[18] Kaplow and Shavell provided a formal proof that even in asymmetrical situations, fairness necessarily dictates a Pareto-inferior outcome in at least some situations (p. 53 n.75). The proof requires only that fairness be a continuous function, so that "[t]he fairness assessment cannot change infinitely at the margin in response to a small, finite change in the level of some consumption good" (p. 54 n.75). The heart of the proof is as follows: if some nonwelfare factor $X$ matters, then at least it must be able to break ties between two states of the world that have equal welfare. If we assume continuity, then the state of the world with the higher $X$ value should still be favored when two states of the world are very close in terms of welfare but not tied. This means that we can end up favoring a state of the world that is Pareto-inferior.

Kaplow and Shavell's theorem could be less dramatically described as "The Impossibility of the Paretian Tie-Breaker." What appear to be innocuous assumptions in a proof sometimes turn out to be stronger and more contestable,[19] but Kaplow and Shavell's formal proof is at least plausible on its face. In any event, I am in no position to quarrel with this reasoning, having endorsed a looser version of this argument myself some years ago in the context of environmental ethics.[20] My

---

18. Louis Kaplow & Steven Shavell, *Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle*, 109 J. POL. ECON. 281 (2001). One of the elegant points of the proof is that it requires only very weak assumptions about continuity and the form of the utility functions.

19. For a critique of the continuity assumption, see Chang, *supra* note 8. A more fundamental objection may be that Kaplow and Shavell's model views fairness as a component of the social goal rather than as a constraint. At least some fairness advocates would reject that conceptualization. *See* Robert Nozick, *Side Constraints, in* CONSEQUENTIALISM AND ITS CRITICS 134, 136-38 (Samuel Scheffler ed., 1988).

20. *See* Daniel A. Farber, *From Plastic Trees to Arrow's Theorem*, 1986 U. ILL. L. REV. 337.

normative conclusion, however, was the opposite — that if we attach intrinsic value to the environment, we should be prepared to make at least minor incursions on the Pareto principle.[21]

Somewhat surprisingly, Kaplow and Shavell do not rely in their book on their formal proof, which they relegate (in paraphrased "heuristic" form) to a footnote (pp. 53-54 n.75). Instead, they rely on more traditional normative analysis. Their normative argument may be clearer if we consider their analysis of tort law, one of several contexts in which they apply their argument.

. Kaplow and Shavell begin by considering reciprocal accidents, where everyone is an accident victim as often as an injurer. The efficient rule is either strict liability or negligence, depending on the legal costs associated with each rule. Strict liability could result in more lawsuits, but they will be individually cheaper, since it is not necessary to consider the issue of fault. There are no distributive consequences since everyone will be a defendant as often as a plaintiff. Thus, the choice of rules affects everyone in exactly the same way. Some notions of fairness might favor one of the two rules. For instance, we might take the view that imposing liability on blameless injurers is unfair. But this fairness-dictated rule might have higher administrative costs than strict liability, making everyone worse off (pp. 103-06). If we are serious about fairness, however, surely we must be willing to accept at least a tiny additional administrative cost to achieve a higher level of fairness. But the administrative costs will be paid by everyone (since everyone is identically situated), so the net result is that everyone is worse off under the fair rule. Thus, unless we are willing to adopt a policy that makes everyone worse off, we must reject the fairness-based rule.

Kaplow and Shavell argue that, to be consistent, we must therefore reject the fairness rule across the board, even in nonreciprocal situations.[22] They supplement this analysis, however, with an independent analysis of the nonreciprocal case, where one group is composed of injurers and the other of victims. Even in the nonreciprocal situation, Kaplow and Shavell contend, the fundamental conclusion is unchanged:

> The relevant point is that, however many individuals might benefit from a fairer rule, and however great their benefit might be, the fact that social welfare is lower means that a judgment has been made that the losses borne by those who are worse off under the rule are of greater social im-

---

21. *See id.* at 346 n.22 ("[E]nvironmental values can at least break ties. Arguing from continuity, however, a strong environmental preference [by the decisionmaker] should at least overcome a slight disparity in efficiency.").

22. Pp. 110-11. The demand for consistency here seems attenuated. Why shouldn't we consider fairness concerns to be triggered only by unequal treatment?

portance with regard to consideration of different individuals' levels of well-being. (p. 120)

Thus, they conclude, except to the extent that ideas about fairness might be factored into the social welfare function, fairness should be irrelevant.

Kaplow and Shavell discuss the practical implication of these findings at some length. As to ordinary citizens, they say, the use of social norms in everyday life remains sensible. People cannot be expected to base their ordinary conduct on complex utility calculations (pp. 382-83). For government officials, however, welfare maximization should be the goal within the constraints allowed by the public's naive belief in social norms of fairness (pp. 396-402). The lesson is even clearer for legal academics and other policy analysts. Kaplow and Shavell maintain that serious policy analysts have only one responsible option: "The duty of analysts is to teach, write, and speak publicly about the virtues of sound analysis, even at the expense of their own popularity, particularly in cases in which such analysis may strike a less expert audience as troubling because its rationale is counterintuitive" (p. 399). In particular, they maintain:

> [w]e do not discharge our responsibility as legal policy experts if we allow ourselves to be guided by simple norms designed for everyday life, if we restrict the sophistication of our academic analysis to ensure that it is fully accessible to those with little welfare economic expertise, and if we resist being viewed as engaged in a sometimes technical enterprise. (p. 390)

And "undertaking welfare economic analysis in an explicit manner ... is the best way to determine which legal rules in fact are most likely to improve the well-being of members of society" (p. 390).

In sum, Kaplow and Shavell claim to have established that "legal policy analysis should rely exclusively on welfare economics" (p. 468). Through pure "deductive logic," they maintain, they have shown that fairness is untenable: "[L]ogical consistency implies" that if one endorses fairness "one has thereby endorsed the view that adopting a legal rule that makes everyone worse off may well be good" (p. 468). Welfare economics is the one and only true path to good policy: "[W]e believe that responsible government decisionmakers will be able to make better policy decisions if those who analyze legal policy devote themselves to identifying the effects of legal rules on individuals' well-being — that is, if they employ welfare economics rather than base their analysis on notions of fairness" (p. 472).

This ambitious program for policy analysts assumes not only that fairness can be rejected as a valid value, but that social welfare functions can provide a complete method for social choice. Since welfarism, a la Kaplow and Shavell, requires the application of a social welfare function to individual utilities, we are immediately faced with two questions. First, assuming the utilities are given, how do we find

the right SWF? Second, how do we get the utilities in the first place? To these we can add a third question: If the welfarist program turns out not to be viable, can economics make any other contribution to identifying equitable outcomes?

## II. Defining Social Welfare

Aggregation is not a problem under the cost-benefit approach favored by Posner. We simply add up everyone's monetary valuations of various states of the world, using existing prices. The option with the highest number is the cost-benefit winner. This procedure has some drawbacks. As mentioned earlier, it is sometimes indeterminate where large wealth shifts are involved. Also, determining the appropriate dollar values is not always easy. Finally, whether the results have anything to say about fairness is debatable.[23] Thus, cost-benefit analysis is no panacea. But for the believer in economic efficiency, the aggregation part of the process could not be easier.

Not so for the welfarist. The concept of a SWF is quite broad in scope: *any* function of utility distributions that increases whenever individual utility goes up. (This assumes that we have already determined utilities, the subject of the next section. For now, however, let us put aside the problem of defining utility and assume that suitable utility functions exist.) Choosing a SWF might be only a minor technical step if we generally obtained the same outcome regardless of our specific choice for the SWF. But in reality, the choice of a specific SWF often controls the outcome. In this Part, I demonstrate that fact and its impact on our understanding of welfarism.

Kaplow and Shavell's theory is incomplete and in some respects ambiguous about the definition of the SWF. To be a serious program, welfarism would need criteria for choosing between SWFs, a specification of which SWFs are candidates for adoption, and a set of non-welfarist constitutional rules for society to follow in choosing between SWFs. Kaplow and Shavell provide none of this. Without such restrictions, however, they cannot exclude the possible decision rule: "Pick the SWF which, when applied, produces the *fairest* possible outcome."[24] Of course, this rule is logically consistent with welfarism, because we will ultimately use only this SWF to design ordinary rules of contract law, criminal law, and so forth. So we can pretend that all

---

23. The best fairness argument is that ex ante, everyone can expect to win from economically efficient rules: the fact that the rules will turn out ex post to favor some and not others should not be a decisive objection when the gamble is one that everyone would have accepted initially. This argument for fairness depends critically on the shift to the ex ante perspective. The validity of this shift turns out to be more complicated and limited in scope than most advocates seem to realize, so the fairness of Kaldor-Hicks remains debatable. For a fuller discussion, see Farber, *Ex Ante Perspective, supra* note 3, for a fuller discussion.

24. More specifically, the fairest Pareto-efficient outcome.

of these legal rules are designed on the basis of the most unalloyed focus on social welfare. But the game will be rigged — we will have chosen the SWF so that the end result of our supposedly undivided attention to welfare will satisfy our notions of fairness. Kaplow and Shavell may have in mind some set of criteria that excludes such SWFs, but they do not provide them.

In any event, Kaplow and Shavell's official position is that no basis exists for selecting a unique correct SWF. As we will see in more detail below, this position implies the existence of nonwelfarist legal rules. If there are at least two possible candidates for the SWF, then society will need a mechanism for choosing between them. (Of course, people would be unlikely to debate and vote on the mathematical formulation of the function: what they would discuss is which criteria should determine the distribution of well-being among the members of society.) Except in the extremely unlikely event that a single mechanism maximizes every individual's welfare, the design of this mechanism cannot be based on maximizing social welfare. Thus, if there are even two candidate functions, Kaplow and Shavell's claim fails: there are at least *some* legal rules (the constitutional procedures for selecting a welfare function) which are *not* based solely on maximizing social welfare.[25]

As mentioned earlier, Kaplow and Shavell may fail to appreciate this problem because (contrary to their official position on this point) they may implicitly assume that the SWF is utilitarian, so there is really nothing to choose. Whether or not this really is their true inclination, it would at least allow them to escape from the need for nonwelfare-based constitutional rules. But it would only exacerbate the problems discussed in Part III.

## A. *The Pivotal Importance of SWF Selection*

As it turns out, the choice of a SWF is just about as outcome determinative as it possibly could be. Given any listing of social outcomes that does not violate the Pareto principle, we can find a SWF that comes out the same way. As shown in the Appendix, we can construct a suitable SWF for any ordering over any finite set of points, so long as the ordering is not chosen in a way that violates Pareto (or that precludes symmetry or convexity if we also desire those qualities). In other words, the concept of a SWF is extremely flexible. As a matter of methodology, this is a desirable quality of the SWF concept

---

25. Moreover, in societies such as ours, the constitutional rules are not necessarily severable from other legal rules. Because the wealthy enjoy more of a political voice in our society, income distribution is, in part, a constitutional issue, not just a welfarist issue. That is, a complete specification of our society's rules for making political decisions (including the decision of how egalitarian to be) would include the rules for obtaining wealth inasmuch as those rules also help determine political influence.

because it allows the use of a SWF to represent as broad a range of orderings as possible. But it also makes the choice of a SWF critical to the results of any welfare analysis. Subject to these restrictions, welfarism is like an Enron accountant: it can give us whatever answer we prefer about how to rank any finite set of alternatives.

This result is stronger in one respect than we need because we don't really need to know the complete ranking of any list of social states, only the social state that is picked as best. It is also weaker than we want because we might have an infinite number of choices. But there is a standard theorem in welfare economics that serves equally well (though a little less dramatically) to show why the choice of a SWF is so critical. According to this theorem, we can always construct a social welfare function which favors whatever Pareto-efficient alternative we happen to like best.[26] In short, as a leading text puts it: "[A] competitive market system will give efficient allocations but it says nothing about distribution. The choice of distribution of income is the same as the choice of a reallocation of endowments, and this in turn is equivalent to choosing a particular welfare function."[27] Thus, adopting a welfarist approach does nothing in itself to solve the problem of social equity. Everything depends on the choice of SWF, and with the right choice of SWF we can justify practically any outcome we want.

This is perhaps an overly involved lead-up to a very simple point. If a welfarist approach means little without the choice of a SWF, and if there are multiple viable candidates, then a choice must be made. That being so, there must be some procedure for making the choice. Obviously, we cannot use welfarism to select the procedure itself: until we have used the procedure, we can't pick a SWF, and without the SWF, we can't practice welfarism. Moreover, it will not normally be the case that everyone in society agrees on a SWF, even if we can get them to put aside the fact that the choice of a SWF will determine the choice of legal rules and thereby their own welfare. So we must choose a SWF-selection procedure on some basis other than welfare. In Shavell and Kaplow's terms, the value used to select the procedure must be a form of fairness, since it is nonwelfarist.

---

26. More specifically, this theorem holds that for any Pareto-efficient allocation, $x^*$, and any concave, continuous, and monotonic utility functions, $x^*$ maximizes some SWF (actually, a weighted sum of the utility functions.) *See* HAL VARIAN, MICROECONOMIC ANALYSIS 198-202 (2d ed. 1984). The proof is quite elegant. If $x^*$ is Pareto-efficient given the resource constraint, we know from the second theorem of welfare economics that $x^*$ is the competitive equilibrium for some initial allocation of resources. That is, there is a set of prices at which supply and demand would balance for each good at $x^*$, starting from that initial allocation. We can then set the relevant weights as the reciprocals of the marginal utility of income for each agent.

27. *Id.* at 209.

This means that, unless we can eliminate all but one candidate SWF a priori, the rules used by society to select its SWF must be determined on some basis other than social welfare. This is not a mere quibble. The question of how to choose a SWF — how to weigh utility — has potentially decisive societal impact. Either everyone in society must get a vote on the result or some subgroup must make the decision. In essence, the rules of constitutional law determine who gets to make the decision and what procedures they must use. The rules must dictate who gets to vote and what discussions are allowed beforehand. Hence, we have at least one important counter-example to the thesis that Shavell and Kaplow propound, which is that fairness is never relevant to the choice of legal rules. At the very least, it seems, fairness must be critical to some of the rules of constitutional law.

Nor can we assume that the social welfare function is chosen once and for all at some kind of constitutional convention. Every time society faces a new policy decision, it might conceivably want to reconsider its choice of SWF. But since there is always a SWF that favors any Pareto-efficient allocation, this simply throws us back on the ultimate, nonwelfarist issue of fairness in every case. Rational individuals will choose a SWF that produces the outcome they most prefer, so nonwelfarist preferences about outcomes will drive the choice of SWF. In short, welfarism may turn out to be vacuous, providing the following recipe for decisionmakers:

1. Choose the fairest of the Pareto-efficient outcomes.

2. Generate a SWF that is maximized at that outcome.

3. Apply the SWF.

4. Pick the predetermined outcome.

## B. *The Utilitarian Solution*

The need to use fairness considerations to select a SWF seemingly ought to be a ground for very serious concern for Kaplow and Shavell. Yet, they seem to regard the choice of a utility function as a purely technical issue, which they mostly relegate to scattered footnotes. Indeed, although they do not say so, they seem to assume that the welfare function will take a single particular form. As they explain their view of the proper procedure, first an analyst chooses a numerical representation of each individual's utility functions and then she chooses a function, F, to "aggregate" these numbers (p. 24 n.15). As they explain, this means that if the initial choice of utility functions is not precisely comparable, we can use the SWF to make adjustments so that the aggregate comes out right. Their focus on "aggregation" is the key. Although Kaplow and Shavell never say so,

their argument really only makes sense if we assume that the SWF is utilitarian and simply adds up individual utilities.[28]

   This implicit assumption about the form of the social welfare function is also crucial to another part of their argument. One of their major criticisms of fairness is that it is applied ex post rather than ex ante (pp. 442-43). But welfarism allows unambiguous reliance on ex ante evaluations only for a narrow class of SWFs. We can illustrate this with a simple example. Suppose that Ann and Mary have identical utility functions which are a linear function of their wealth. (Hence, they are risk neutral.) Suppose that, for each of them, U($1,000) = 1 and U($0) = 0. Now they get a chance to play a lottery for the price of $1000 apiece. If they agree to play, a coin will be flipped, and the winner will get $3,000. The expected value of the lottery to each is ½ of $3,000, or $1,500, which gives them a utility of 1.5 units as opposed to the 1 unit they currently enjoy. Hence, they will both want to play.

   Now, imagine that society is faced with the decision of whether to allow such a lottery. If the SWF simply adds utility functions, then it doesn't matter whether we determine social welfare ex ante or ex post. If we do so ex ante, we add the two expected utility figures of 1.5 to obtain a total social welfare of 3. If we perform the same calculation ex post, we know that one of the individuals will have a utility of 0 and the other will have 3, for the same total as before. Hence, social welfare is the same, whether we evaluate it ex ante or ex post.

   Suppose instead that society prefers utility to be evenly distributed. For example, instead of adding individual utilities, the SWF might add the square roots of those utilities, a calculation which favors more equal distributions. Using this SWF, society's current welfare level is the square root of one plus the square root of one, or 2. Should society agree to the lottery? If we perform the calculation ex ante, we add the square root of 1.5 to the square root of 1.5. This sum is greater than 2, the welfare level of the status quo. Hence, we should approve the lottery. But if we perform the calculation ex post, then the welfare level is determined to be the square root of 3 (the winner's utility) plus the square root of zero (the loser's utility). This is less than 2, the welfare level of the status quo, so the ex post analysis says to reject the lottery. With this welfare function, although the lottery would obtain everyone's consent ex ante, it turns out to be welfare-decreasing ex post. In other words, the welfarist tells us to approve a lottery ex ante which is known with complete certainty to make society worse off ex post. As Kaplow and Shavell themselves observe in passing in a

---

28. Indeed, at one point, they use the telltale phrase: "once a social welfare function is chosen (that is, once a judgment about aggregation is made[)]. . . ." P. 27 n.23 (internal cross-reference omitted). The word "aggregation" suggests that they are implicitly assuming that the SWF is additive.

footnote, the only way to avoid such discrepancies and maintain the purity of the ex ante approach is to use a utilitarian SWF (p. 443 n.95).

This problem also affects at least one of Kaplow and Shavell's specific policy recommendations, which involves criminal punishments. If detecting and punishing an individual offender is costly (as it always is), we can save money by punishing fewer offenders but ramping up the sentence to reflect the decreased probability of detection. Thus, if the harm is 100 and there are four offenders, we can get equivalent deterrence either by catching all of them and imposing a sentence of 100 or by randomly catching one of the four and imposing a sentence of 400 (pp. 319-20). Given the enforcement costs, we should adopt the latter alternative, contrary to the usual view that sentences should be proportional to the harm in individual causes.[29]

If society does not use a linear welfare function, however, it may prefer to impose a punishment on all convicted criminals at level 100 despite the higher enforcement costs, rather than punishing a quarter of the criminals at level 400. We can see this effect with our two-person society, defining the SWF as before to be the sum of the square roots of the individual utilities. Suppose that our two individual criminals are each currently at a utility level of 2. We have a choice between randomly catching one of them and imposing a punishment of 2, or catching both and imposing a punishment of 1. Ex ante, both the individuals and society are indifferent between these alternatives, so even minute enforcement costs would be enough to favor adoption of the random strategy. But ex post, society would regret this choice as inequitable. With uniform fair punishment, both criminals end up with utilities of 1, leading to an ex post social welfare level of 2. With random heightened punishment, one criminal ends up at 2 and the

---

29. For a discussion of the more conventional view, see BIX, *supra* note 3, at 110-11. The form in which Kaplow and Shavell cast their argument does not seem to be quite right. They argue that this alternative (less likely but more serious punishment) is perfectly fair because when the expected value of the punishment is greater than the benefit of committing the crime, deterrence is completely effective and no one ever has to be punished. P. 319. "Note as well," they argue, "that the only real sense in which potential criminals are treated poorly when the sanction is 400 is that they are deterred from committing a crime that they would find it in their self-interest to commit were they under the regime with the fair punishment." P. 324. But zero crime is not a Nash equilibrium, because society's threat to continue to invest in crime detection when the crime rate is zero is not credible. (For a discussion of this equilibrium concept, see MAS-COLELL ET AL., *supra* note 13, at 246-50.) In their basic model, Kaplow and Shavell assume this issue away by positing a fixed, costless level of crime detection. P. 318 n.48. This is clearly unrealistic. If no crimes are ever committed, a rational society should invest no resources in crime detection. But when that happens, no criminals are punished, so violations increase. The solution is clearly a mixed equilibrium in which society sometimes enforces and criminals sometimes offend. At any level of penalty, some criminals will still violate the law and some of them will be punished, with the latter proportion depending on how much society invests in law enforcement. But this does not affect the basic argument that Kaplow and Shavell want to make, which is that we can get the same amount of deterrence with lower transactions costs by using a higher level of punishment. As they correctly observe, their basic claim holds even with imperfect deterrence. Pp. 331-36.

other at 0, leading to an ex post social welfare level equal to the square root of 2. This is lower than the ex post welfare level for uniform penalties. Thus, ex post, society prefers the more equal level of punishment. A welfarist can confidently advocate the "efficient punishment" scheme only by assuming that the ex post and ex ante welfare levels will be identical. And this will be true only for a utilitarian SWF.

As we have seen, some aspects of the Kaplow and Shavell argument makes sense only if we assume that they are not merely welfarists but utilitarians, so there is essentially only one possible SWF. Thus utilitarianism ensures that we can use ex ante evaluations without trepidation and also that we do not have to worry about fairness-based constitutional rules for picking the SWF. In short, Kaplow and Shavell provide an important argument in favor of utilitarianism. Essentially, they have shown that if we assume the possibility of a interpersonal comparisons of utility and demand that ex post and ex ante social judgments agree, then only utilitarianism is consistent with the Pareto principle. (In principle, a utilitarian SWF would allow different weightings for different individuals' utilities. But if we do the utilities right in the first place, then Kaplow and Shavell's symmetry principle requires equal weightings, since we are supposed to count the properly computed measure of each person's utility equally.) The converse of the conclusion is equally significant: if we do not demand this form of temporal consistency, there will *not* be a unique SWF, and some fairness value must be invoked as a basis for the constitutional rules used by society to pick a SWF.

The demand for temporal consistency is not an unreasonable one, so utilitarianism enjoys a significant advantage under this analysis.[30] But utilitarianism puts additional pressure on other parts of their theory. Other SWFs pay some attention to how utility is distributed. For example, the square-root function discussed earlier favors equal distribution of utility, which in normative terms means that we attach independent importance to having every individual enjoy a reasonably good life. Utilitarianism is a more radical break with individualism,

---

30. This may be an even stronger argument for utilitarianism than the discussion in the text suggests. When we talk about ex post and ex ante analysis in this setting, we are not really talking about the passage of time. Rather, we are talking about a change in our information about the impact of a decision — but the additional information seems to be irrelevant! In the examples discussed earlier, we know in the ex ante situation that one of two individuals will be affected; in the ex post situation, we know the name of the affected individual. But since the individuals are identical in every relevant respect, it's hard to see why knowing or not knowing the name should change society's choice; after all, we knew that it had to be one of them or the other anyway. Thus, requiring consistency between the ex ante and ex post perspectives really just means that the presence or absence of irrelevant information should not affect societal preferences. This seems like an eminently reasonable requirement. If we accept this relevance requirement, we get equivalence between ex ante and ex post assessments, and hence the SWF must be utilitarian.

since it treats individuals as nothing more than vessels into which utility can be poured. In theory, the utilitarian would be happy to have the entire human race extinct except for one person, if that one person could be made ecstatic enough as a result. Thus, utilitarianism raises more troubling ethical concerns than other forms of welfarism.

Moreover, utilitarianism also raises the ante in technical terms by requiring unit-for-unit comparisons between different individuals' utilities. The highest possible demands are therefore made on our ability to make interpersonal comparisons. Apart from the technical problems, Kaplow and Shavell face the challenge of defining utility in such convincing terms that we are willing to accept the Pareto principle as an absolute dogma regardless of any other moral principles. Thus, Kaplow and Shavell can escape the problems discussed in this section through utilitarianism, but only at the cost of worsening the problems discussed in the next section.

## III. DEFINING UTILITY

There are three main issues relating to utility. Exactly what preferences count in determining utility? Are preferences an appropriate basis for determining social welfare? And can we actually make the required measurements of utility as a practical matter? Since utilitarianism has been around for well over a century, there is a voluminous literature about these issues. Fortunately, we need not consider the issues in great depth. It is clear that at best welfarism is a long way from being a practical method of deciding all questions of equity. Of course, we could easily create *some* scheme of utility indexes. But it is harder, by orders of magnitude, to create an index that guarantees aggregate social utility really will indicate the overall well-being of everyone in society in some meaningful sense.[31]

We begin with the first question: defining the relevant set of preferences. In terms of the types of preferences that count, Kaplow and Shavell take a forthright position. According to them, all preferences are created equal.[32] We should not, therefore, discard preferences such as racism or sadism in calculating social welfare:

> To trump preferences is, in essence, to redefine individuals' well-being in a manner that substitutes some other preferences — ones that are cleansed, so to speak — for individuals' actual preferences. . . . But such an approach is troubling from the perspective of welfare economics because the moral force and appeal of welfare economics lies in promoting

---

31. For insightful general discussions of these issues, see KENNETH J. ARROW, SOCIAL CHOICE AND INDIVIDUAL VALUES 109-18 (2d ed. 1963); AMARTYA SEN, ON ETHICS AND ECONOMICS 30-57 (1988).

32. It should be noted that the other most prominent welfarist in the legal academy takes the contrary position. *See* Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FLA. ST. U. L. REV. 241, 262-67 (2000).

the actual well-being of people, not in advancing some hypothetical no-
tion of satisfaction that is distinct from that of the individuals who are the
objects of our concern. Furthermore, employing a cleansed version of
preferences rather than actual preferences may lead one to favor policies
that make everyone worse off, just as when a notion of fairness is
pursued at the expense of individuals' well-being. (pp. 419-20)

Moreover, they add, "[t]he idea of an analyst substituting his or
her own conception of what individuals should value for the actual
views of the individuals themselves conflicts with individuals' basic
autonomy and freedom" (pp. 421-22).

There are two problems with this claim. The more obvious one is
that it makes the link between utility and "well-being" problematic.
Although we may want the lives of all individuals to "go well," it is not
clear that this means helping them to satisfy *any* possible preference
they might have, no matter how morally objectionable.[33] We may not
think, for example, that satisfying the preferences of a rapist should
count as increasing his well-being or should prima facie be counted as
an improvement in social welfare (from which his victim's pain will
then be subtracted.). If Shavell and Kaplow did a global search and
replace, substituting "desire gratification" for the words "welfare" and
"well-being" throughout their book, the meaning would be entirely
preserved but much of their normative argument would fall flat. For
example, their basic theorem would read: "Adopting fairness as a
binding norm necessarily entails the occasional possibility of at least
slightly reducing the desire gratification of everyone in society." It is
not self-evident that the rephrased version is a knock-down argument
against fairness.

Kaplow and Shavell attempt to answer this objection, but their
answer seems to come down to the empirical claim that in most
instances, objectionable preferences will be outweighed by the bad
consequences to others of fulfilling those preferences (pp. 427-31).
They do not really dispute that objectionable preferences may win out
under their approach in some cases. Counting fulfillment of these
objectionable preferences is necessary to their version of the Pareto
principle. But their own argument against fairness can be turned
against them here. They argue that fairness must be rejected as a
general principle as a matter of consistency, because in some
circumstances fairness violates our commitment to the Pareto
principle (p. 56). But now we can equally argue that the welfarism
(and with it, the Pareto principle) must be rejected as a matter of

---

33. Alternative and perhaps more appealing views of utility have been offered by other
utilitarians. *See* J.J.C. SMART & BERNARD WILLIAMS, UTILITARIANISM: FOR AND
AGAINST 12-28 (1973).

consistency, because in some circumstances it violates our commit-
ment to other moral principles.[34]

Moreover, the Kaplow-and-Shavell view of utility, rather than
respecting individual values as an aspect of autonomy and freedom,
actually implies that society should be free to manipulate individual
desires at will. Kaplow and Shavell view the failure to anticipate future
preference shifts as being a form of imperfect information, which
society should correct (p. 411). In their view, society can treat such a
change in preferences either as an individual's failure to anticipate his
own future preferences or as a failure to anticipate how his fixed
preferences would respond to much different circumstances (p. 412
n.31). Hence,

> if it is discovered that individuals are able to adapt to certain physical
> disabilities more or less readily than is commonly supposed, the valua-
> tions employed in measuring tort damages or in performing cost-benefit
> analysis (for example, of highway safety improvements) should reflect
> actual harm rather than victims' uninformed ex ante estimates. (p. 413)

The logical implication seems to be that, if accident victims perfectly
adapted their preferences to their new limitations, we should not
count a serious disability as a utility cost to them at all. Prospective
victims currently might think that they prefer not to be disabled, but
we (the exalted legal policy analysts) would know better.

Again, whatever might be said for this view of well-being in the
abstract, it does make the Pareto principle less than a self-evident
moral truism. Suppose, for example, that a change in tort law saves a
penny in everyone's insurance bill, while adding only one accident
victim. Assume that the accident victim, while severely disabled, will
adapt psychologically to the disability and in fact will be happier as a
result because of his access to powerful painkillers. In this situation,
fairness might require us to veto what Shavell and Kaplow regard as a
Pareto improvement, but so what?

In sum, it is not clear that utility, as Kaplow and Shavell define it,
can actually carry the moral weight necessary to uphold their
argument against fairness. Reducing the Pareto principle to the Prozac
principle deprives it of some of its moral force. It is possible that the
definition of utility could be refined so as to make the Pareto principle
more compelling, but whether this can be done without causing other
difficulties for welfarism remains to be seen.

So far, we have been considering the difficulty of defining
individual utility in a way that convincingly corresponds to some

---

34. As a fallback, they suggest that preference-cleansing must be rejected because of the
absence of any clear standards for doing so. Pp. 418-31. But again, this argument can be
turned against them. Because it fails to specify clear grounds for choosing a SWF, their own
theory would fall prey to the same demand for clarity that they impose on others.

notion of "well-being."[35] Yet welfarism requires not only that we do this on an individual basis, but that we be able to make interpersonal comparisons of utility and in some way combine them. Whether combining individual utilities is morally appropriate or even possible is highly controversial. Here, again, there are two subissues.

The first subissue is whether combining individual utilities is appropriate even if it is possible. Posner makes the classic argument:

> [B]y aggregating utility across persons, utilitarianism treats people as cells in the overall social organism rather than as individuals. This is the source of the familiar barbarism of utilitarian ethics, such as the deliberate sacrifice of innocents to maximize the total amount of happiness in the society (or in the world, or the universe), or the "utility monster" whose capacity for sadistic pleasure so far exceeds the capacity of his victims to experience pain that utility is maximized by allowing him to commit rape and murder. Defenders of utilitarianism seek to deflect such criticisms by pointing out that lack of trust in officials would defeat any effort to empower the state to attempt to maximize utility on an individual basis. . . . But practical objections to the logical implications of utilitarianism miss the point. The logic itself is repulsive. Even if we assume away all the problems of implementation, and contemplate the result — the inducement of blissful trances by utterly benign, democratically responsive officials — we still don't like it.[36]

Thus, maximizing the value of the SWF may not be an appealing way to formulate society's goals.

The second argument against utilitarianism, which applies more generally to welfarism, is that quantitative comparison of individuals' utility is meaningless. Despite philosophical qualms, most of us have no hesitancy in concluding on occasion that one individual's well-being is higher than another's. But welfarism requires that we be able to make such judgments with great precision between any two hypothetical individuals. For instance, if Abby likes apples and Paula likes poetry, we need to be able to say how many apples Abby needs to eat in order to obtain the same degree of satisfaction that Paula gets from reading each specific poem. Partly this can be seen as a measurement problem, but partly it is a philosophical one: Does it really mean anything to try to reduce these two very different experiences to a common metric?

If all we have to go on is the preference listings of each of the two individuals, we may find it impossible to come up with a cardinal scale. We know that Abby likes having one apple, likes having two better, etc. We have similar information about Paula. We could obviously

---

35. This has been the primary ground for criticism of Kaplow and Shavell to date; all of the sources in note 8, *supra*, at least touch on (some are devoted exclusively to) this issue. *See also* Herbert Hovenkamp, *The Limits of Preference-Based Legal Policy*, 89 NW. U. L. REV. 4 (1994).

36. Posner, *supra* note 1(article manuscript at 11-12).

assign common numbers to the scales — for example, assigning a utility of one to each apple for Abby and each poem for Paula. But it is very unclear whether, even in principle, there is a convincing way of converting their individual rankings into a shared numerical scale.[37] The alternative is to use some gauge of the intensity of preferences (as opposed to their rankings). This seems at least theoretically possible. For example, it might turn out that our common sense judgments about the relative happiness of different people correlates closely with their levels of certain neurotransmitters. If so, we could use the neurotransmitter levels as our utility index. (Under this approach the welfarist motto would be: "[T]he greatest endorphins for the greatest number.") Whether this is a morally appealing goal would remain unclear, but at least it would operationalize the concept of interpersonal utility comparisons. Of course, utilitarians have a variety of other ingenious ideas about how to measure well-being, and one of these may eventually work. At present, however, basing a theory on quantitative comparisons of welfare is reminiscent of the punchline of the famous joke, in which an economist solves the problem of how to open a can on a desert island simply by proclaiming, "Assume a can opener."

Even if we could reliably compare the welfare levels of different individuals, we would still be left with the third problem — we would need to determine each individual's full set of preferences, so we can connect utility states with outcomes. The difficulty is that, if they know the information will be used to determine social policies that affect them, individuals have an incentive to misrepresent those preferences. It turns out to be very difficult to design schemes that will provide the proper incentives for individuals to truthfully reveal their preferences. Such schemes suffer from two general flaws. Such schemes can become quite complex, often making them unworkable in practice. In addition, they have an unavoidable fault, which is that they do not operate on a balanced budget. In other words, the schemes require taxes and side payments that do not even out, so either society must provide subsidies to the participants or must expropriate some of their assets.[38] Of course, the cost of calculating and then utilizing complete information about individual preferences may seem like a mere technical detail. But information problems are basic to modern economic analysis. We should bear in mind that in the absence of similar kinds

---

37. Daniel Hausman argues that the only logically coherent method of doing so is to assign a utility of zero to the very worst possibility that each individual can imagine and a utility of one to the best. This has the unfortunate effect, however, of making the social welfare function depend on how imaginative individual members of society happen to be, which seems implausible. *See* Daniel M. Hausman, *The Impossibility of Interpersonal Utility Comparisons*, 104 MIND 473, 480-81 (1995).

38. *See* VARIAN, *supra* note 26, at 256-59.

of informational costs, a completely centrally planned economy could perform exactly as well as a free market. While it is theoretically possible to design a mechanism to force truthful revelation of preferences, the practical barriers are severe.

At the extremes, the direction of a change in collective welfare may be clear for *any* plausible method of defining and measuring welfare, allowing the welfarist to come to reasonably definite conclusions. But if welfarism is to play the kind of role envisioned by Kaplow and Shavell — completely replacing all other norms for social decisionmaking — we need more than the unproven promise that the utility-measurement problem can be solved. Certainly no such solution is on the horizon. In the meantime, however, it is worth considering whether economics can make any contribution to addressing equity problems on a less global scale.

## IV. IMPLEMENTING EQUITY

In this Part, I will provide a brief survey of some of the ways in which economic theory actually can contribute to our understanding of equity issues.[39] The methods discussed below in Section IV.A have been extensively discussed by tax scholars. In contrast, those discussed in Sections IV.B and IV.C cry out for further elaboration. That will have to wait, however, for another occasion.

### A. *Optimal Taxation*

Prior to their article on welfare and fairness, Kaplow and Shavell were best known for their double-distortion argument. This argument holds as a general matter that the tax system is a more efficient way of engaging in redistribution than the regulatory system. Although their full argument is technical and complex, the basic rationale is straightforward. Suppose we readjust some economically efficient legal rule so that the new rule redistributes income from the rich to the poor. Since we have moved away from the efficient rule, there is necessarily a direct efficiency cost to the rule change. But there is also a second distortion. Just as someone considering an additional hour of work might be deterred by an increase in the marginal tax rate, so they might be deterred by knowing that they are making themselves targets for disadvantageous legal rules. Thus, besides the direct distortionary effects of the redistributive rule, it also has the same distortionary effect on the labor supply as the income tax. Hence, Kaplow and

---

39. Empirical economics may also have a useful contribution to make. Empirical findings provide considerable evidence about the structure of fairness norms in our society. *See, e.g.,* Norman J. Finkel, *But It's Not Fair!,* 6 PSYCHOL. PUB. POL'Y & L. 898 (2000); Martin Nowak et al., *Fairness versus Reason in the Ultimatum Game,* 289 SCIENCE 1773 (2000).

Shavell maintain, we can always transfer the same amount of funds more efficiently just by raising the tax rate.[40]

This conclusion is not ironclad, as Kaplow and Shavell themselves admit.[41] If what we are trying to equalize is not income but rather some quantity that correlates with income, or if individuals differ in their responses to legal incentives, legal rules can add to the efficiency of redistribution.[42] We can illustrate this with the following simple example. Suppose that we are utilitarians and want to maximize the sum of utility (putting aside the definitional and measurement problems discussed in the last section). If income has declining marginal utility for individuals, then redistribution of income from the wealthy to the poor will increase overall utility. But if individual utility functions differ, then at any given income level, the marginal utility of income will be different for some people than for others. We can improve our redistribution scheme by adding a feature that sorts people out according to their marginal utility for money. We might be able to do this with a simple modification in tort damage awards.

Suppose that some type of accident can be avoided simply by being more attentive and sober-minded, but that it is more fun to be carefree. Let's assume that the resulting utility loss is the same for everybody. When people are careless, they sometimes cause accidents. Given those assumptions, we can improve our redistribution system by employing a contributory negligence regime and adding a surcharge to damages. This is by no means a far-fetched hypothetical. Although current tort rules ignore income levels, it's easy to imagine juries augmenting damage awards against wealthy individuals (as revealed, for example, by being the driver of a luxury car involved in an accident).

Here's how the damage surcharge works. In deciding how careful to be, individuals compare the lost utility from being careful with the potential lost utility from paying damages. The first of these items is fixed but the latter depends on the individual's marginal utility for money. Hence, all things being equal, people with a lower marginal utility for money will be more likely to be careless, whereas people with a higher marginal utility will not. Thus, the level of care is a signal of marginal income utility. Now consider a tort suit. The defendant will lose only if he was careless, which means that he is more likely to

---

40. The theory is summarized in Louis Kaplow & Steven Shavell, *Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income*, 29 J. LEGAL STUD. 821 (2000) [hereinafter Kaplow & Shavell, *Should Legal Rules Favor the Poor?*].

41. *See id.* at 827-32.

42. For more complete explanations of this point, see Chris William Sanchirico, *Deconstructing the New Efficiency Rationale*, 86 CORNELL L. REV. 1003 (2001); Chris William Sanchirico, *Taxes Versus Legal Rules as Instruments for Equity: A More Equitable View*, 29 J. LEGAL STUD. 797 (2000).

lose if his marginal utility of income is low. (Recall that people who remain careless despite the prospect of tort litigation have a low marginal utility for money.) Because we are assuming a contributory negligence regime, the plaintiff will win only if he was careful, meaning that he has a high marginal utility of income. Thus, the damage surcharge will transfer money from individuals with a lower marginal utility for money to those with a higher marginal utility, just the kind of redistribution favored by the utilitarian.[43] In contrast, the income tax collector does not seem to have any method of determining individual differences in utility curves, so he cannot take advantage of these difference in marginal utility between individuals who have the same income. In theory, then, the tort surcharge could improve the redistributive accuracy of the legal system. This improved redistribution comes at an efficiency cost (an inefficient decrease in the number of accidents and possibly an indirect effect on labor supply). But the efficiency cost might be worthwhile if we sufficiently value equity.[44]

Thus, the double-distortion argument is not an ironclad proof of the universal superiority of taxes over legal rules in redistribution. Nevertheless, it makes two very important contributions to our understanding of redistributive legal rules. First, the argument probably

---

43. We could achieve a similar effect with a cross subsidy. We could require insurance companies to surcharge (beyond the actuarially fair level) individuals with traffic tickets and correspondingly lower their rates for safer drivers. This would work, however, only with mandatory insurance, since otherwise the less careful drivers would refuse to insure. The idea of using insurance as an alternative to taxation is discussed in Ronen Avraham & Kyle P. Logue, *Redistributing Optimally: Of Tax Rules, Legal Rules, and Insurance*, 56 TAX L. REV. (forthcoming 2003).

44. The labor supply effect exists because in this model, causing accidents amounts to a luxury good — all things being equal, when a person earns more money, her marginal utility for money declines, so damages sting less, so she is more able to be careless and have more fun. Thus, all things being equal, a person with a higher income will find carelessness more affordable. But if most accidents take place during leisure hours, there is a countervailing effect on labor supply, since the utility of leisure declines (through elimination of the fun leisure activity of careless driving) and the relative attractiveness of labor therefore increases. For many luxury goods, we can create the same positive effect on labor supply through an excise tax. We might obtain a similar "tax" on carelessness to some extent by issuing expensive traffic tickets. (The excise tax option is discussed in Kaplow & Shavell, *Should Legal Rules Favor the Poor?*, *supra* note 40, at 825-26.) But here, relying purely on traffic tickets would lose the second benefit of the tort surcharge, which is that the payments are targeted to the least careless individuals. Perhaps, however, we could offer a tax credit to individuals who had gone through the year without a ticket. The efficiency picture for the combination of traffic tickets and tax credits is mixed. This combination of measures avoids the expense of litigating torts cases, and if people are risk-averse but not fully insured, it may have less of an overdeterrence effect than the tort scheme. On the other hand, the cumulative expense of traffic patrols and numerous traffic-court cases might be more than the cost of litigating a much smaller amount of torts cases, the accuracy of the results might be less, and we would have to create a system to cross-tabulate income tax returns and traffic records. In the general case, designing and implementing a tax that duplicates the effects of a redistributive legal rule may require unrealistic amounts of information about each individual. *See* RONEN AVRAHAM ET AL., REVISTING THE ROLE OF LEGAL RULES AND TAX RULES IN INCOME REDISTRIBUTION (John M. Olin Ctr. for Law & Econ., Univ. of Mich.., Working Paper No. 02-004).

holds in many situations at least to a first approximation, meaning that the required redistributive adjustments in legal rules often are small. Second, it tells us where to look for the exceptional cases in which substantial legal redistribution is appropriate.[45] These occur primarily when income provides a signal of some underlying quantity that we want to redistribute, but responses to a legal rule add significant information beyond the income signal. These are nonobvious conclusions, and anyone who is interested in redistribution can benefit from understanding this economic analysis.[46]

## B. *The Nash Bargaining Solution*

Economics may also be able to help us address another equity problem. Suppose two individuals disagree about how to divide a fund between them, and an arbitrator is called upon to resolve the dispute. Assume that we have access to at least some utility information about the parties. What standard should the arbitrator use to divide the money? There are several potential answers, but the one that stands out is the Nash bargaining solution.

There are several plausible characteristics that we might demand from a solution. First, the result should be somewhere between the entitlements of the two parties. After all, the arbitrator is being asked to divide the amount in dispute, not to seize and reallocate amounts that clearly belong to one party or the other. Second, changes in entitlements should only affect the outcome if they are relevant in the sense that the change makes the newly preferred outcome available when it previously would have violated one of the parties' entitlement. Third, the outcome should not depend on how the arbitrator scales the utilities of the two parties. The reason for this requirement is that the best the arbitrator can hope to do by observing their behavior is to establish a von Neumann-Morgenstern utility function for each one, but these functions are unique only up to multiplication. (For this reason, utilitarianism is not an option here.) Fourth, if the parties have identical utility functions in the relevant range, they should split the fund equally, since there is no reason to prefer one party to the other.

Nash proved that these four requirements are compatible only with a single solution. We obtain that solution as follows: take any possible division of the disputed amount; calculate the utility gain of

---

45. Or more accurately, one of the main places to look. It is also possible, for example, that public choice problems might make it easier to use the courts than the legislature for redistribution. Certainly, recent rounds of tax legislation have not been an edifying spectacle.

46. Other important work by tax analysts also uses economic methods to shed light on the best methods of achieving equity. *See* Reuven Avi-Yonah, *Why Tax the Rich? Efficiency, Equity, and Progressive Taxation*, 111 YALE L.J. 1391 (2002) (book review) (summarizing much of the important literature on the subject).

each party compared with their entitlement level; multiply the two utility numbers. The bargaining solution is the point which maximizes this product.[47]

Besides satisfying these plausible assumptions, the Nash solution has another substantial strength. There is a good argument that it represents the outcome that rational bargainers would actually achieve. Essentially, the Nash solution penalizes the party who is the most risk averse and has the most to lose from a breakdown in bargaining, which seems realistic.[48] Consider a bargaining situation where parties with different discount rates exchange offers, knowing that if their offer is rejected, a bargain will be delayed with consequent loss of utility. As we make the time between offers smaller and smaller, the outcome of this bargaining game converges to the Nash solution.[49] Thus, if the parties could conduct their bargaining instantly and with perfect rationality, they would arrive at something like the Nash solution. This makes it a plausible outcome for the arbitrator to pick as fair. Neither party can complain of being cheated of anything that they would have won in a fair negotiation.[50]

Neither the axiomatic approach nor the alternating-offer model provides a knockdown argument for the fairness of the Nash solution. At best, the Nash solution is fair only if the initial bargaining situation itself was equitable. The bargaining situation includes the initial endowments of the parties and their utility curves for the disputed funds (which might be affected by their incomes). Thus, the Nash solution is probably not a very useful tool for considering large questions of distributive justice.[51] But in a number of legal situations, decisionmakers must compare some actual situation with the results of a hypothetical fair bargain. The Nash solution can help put meat on the bones of this concept.

A couple of examples may help illustrate the possible applications. Doctrines such as fraud and duress provide a basis for invalidating a contract, but in a particular case, abusive bargaining may be too difficult to prove or too subtle to fit these doctrines. One way of gauging whether the bargaining process malfunctioned is to see whether the outcome was at least roughly what we would expect from a fair

---

    47. The Nash solution is explained in BRIAN BARRY, THEORIES OF JUSTICE 12-17 (1989).

    48. *Id.* at 23-24.

    49. *See* KEN BINMORE, JUST PLAYING 122-28 (1998).

    50. *See* BARRY, *supra* note 47, at 24-30.

    51. *See* JOHN E. ROEMER, EGALITARIAN PERSPECTIVES: ESSAYS IN PHILOSOPHICAL ECONOMICS 202-06 (1994).

bargaining process.[52] The Nash solution provides no basis for questioning the initial bargaining position and hence cannot be a basis for redistribution in favor of the weaker party. What it may be able to do, however, is at least prevent the use of abusive bargaining to redistribute even more wealth toward the party who is already stronger.

Other examples might arise in tax or bankruptcy law, both of which sometimes require a judgment about whether a contract between related parties is comparable to an arms-length agreement.[53] If the market is fairly thick, it may be possible to make this determination simply by a comparison with market transactions. But where the market is thin, this may not be possible. Rather than simply relying on intuition, the Nash solution might give us a gauge for making this determination.

The biggest question about the Nash solution is not whether it is the perfect definition of bargaining fairness. There is no reason to think it is, but we do not need a perfect definition to address practical legal problems. Rather, the problem is the same as that of welfarism: Can we get the information needed to apply this technique? Fortunately, the Nash solution is not as information intensive as welfarism. We do not need full interpersonal comparability of utility levels, only estimates of attitudes toward risk. Moreover, since we are not purporting to address every issue of social policy, we can avoid situations where utility assignments are especially likely to be idiosyncratic or difficult to determine. In many situations, the bargaining will essentially be over money, the most fungible of commodities. Even so, the Nash solution may be difficult to determine in real-world situations. Still, it has better prospects for success, albeit in a smaller sphere, than the more ambitious program of welfarism.

## C. *Fair Division Games*

Welfarism and the Nash solution both require the analyst to determine the optimum outcome. An alternative approach is to concentrate on devising mechanisms that will guide the parties toward

---

52. *See* MICHAEL J. TREBILCOCK, THE LIMITS OF FREEDOM OF CONTRACT 47-48, 120 (1993) (suggesting that procedural unfairness may be inferred when contract terms differ greatly from the norm).

53. In tax law, the best example is provided by § 482 of the Internal Revenue Code, I.R.C. § 482 (2003), which provides the Treasury with the authority to reallocate income among related entities in accordance with the true substance of the transaction. For example, if two related entities transact, § 482 allows the Treasury to redetermine the transaction in a manner consistent with an arms-length transaction. So, if Distributor (located in the United States) buys goods from related Manufacturer (located in a lower tax jurisdiction), and the purchase price is too high (versus comparable arms-length transactions), the IRS can effectively change the purchase price for tax purposes, increasing Distributor's taxes. The so-called "transfer-pricing agreements" that are frequently entered into by multinationals with the government are essentially agreements (prior to the transaction) on reasonable prices.

a fair outcome. The difficulty is that we normally cannot compel full disclosure of preferences, which the parties have every reason to conceal. The parties would ideally like to maximize their own shares, but this is impossible for all of them to achieve. We cannot impose a solution that maximizes the parties' collective welfare, and they have no reason to make concessions merely to improve someone else's welfare. Thus, welfare-maximization is not a realistic goal here. A more realistic goal is a division that is efficient (so that it could not be improved by voluntary readjustments) and envy-free (so that no one would be willing to exchange their own share for anyone else's).

The classic example is the cake-cutting problem. How can we ensure that two children share a cake equally, when neither is altruistic and both would prefer as much cake as possible? The answer, known to parents for generations, is "cut and choose" — one child cuts the cake in two pieces and the other one chooses. Considerable ingenuity has been employed in generalizing this procedure. Matters rapidly become more complex when the cake is not homogenous. For example, most of the decorations might be on one part of the cake and the children might differ in how much they care for the decorations versus the cake itself. Further complexities abound as the number of players increases. In the end, once we leave behind the simplest situations, the various "multiple moving knives" solutions become increasingly impractical to implement.[54]

Nevertheless, there are some promising though imperfect procedures for real-world situations. A leading example is the adjusted winner ("AW") procedure. This procedure applies when two individuals must divide a finite number of indivisible assets. The two parties are each given 100 points, which they can divide up among the objects. The parties then simultaneously announce their allocation of points to various objects. Each individual is tentatively awarded all of the objects for which she has the "high bid." One of the individuals is given the tie-breaking advantage in this part of the procedure. Generally, this will leave one of the parties with items worth more points to her than the other's total points. To equalize matters, we calculate the ratios for each of the "winning" party's items between the number of points assigned by the winning party and the number of points assigned by the losing party in round one. Beginning with the items with the lowest ratio, we start transferring items to the losing party until both parties have attained an identical point count; this may require selling a single item and dividing the proceeds, or otherwise providing proportional ownership of that item.

The result of the AW procedure, assessed in terms of the announced point valuations, is efficient (the parties have no incentive

---

54. *See* STEVEN J. BRAMS & ALAN D. TAYLOR, FAIR DIVISION: FROM CAKE-CUTTING TO DISPUTE RESOLUTION 6-64 (1996).

to trade), envy free (neither would switch with the other), and equita-
ble (the two end up with equal utilities based on their announced
valuations). The only problem with this procedure is that, at least in
theory, a player with enough information about the other party's
preferences can gain an advantage through strategic allocation of
points in the first round. In practice, however, this type of strategic
manipulation appears to be difficult to accomplish; in any event, there
is another procedure called proportional allocation that can be used as
a backup.[55]

The AW procedure has several possible legal applications. One
obvious application of the adjusted winner procedure is the division of
property in divorce.[56] A similar application would involve the division
of personal property in an estate, where sale to an outsider might fail
to capture the items' sentimental value.[57] In dissolving a partnership,
the procedure might also be used if some of the partnership assets are
more valuable to firm members than to outsiders, assuming that an
auction did not appear to be a fair solution. Compared to the Nash
solution, the AW procedure seems to have a narrower range of poten-
tial applications, but it does have the advantage of being easily
implemented.

## CONCLUSION

As we saw in our discussion of welfarism, one of the biggest
problems in devising equitable solutions is that preferences are private
information, which may be very difficult to obtain. As it happens,
however, economists are interested in the general problem of creating
information-revelation mechanisms. There is good reason to hope for
some useful synergy with law, which could result in some useful new
methods for making equitable allocations.

Economics also has other uses in trying to achieve equity. As we
have seen, it provides useful guidance on the choice of redistributive
instruments. Economic analysis also reveals unexpected connections
between different aspects of ethical judgment — revealing, for
example, the connection between utilitarianism, the Pareto principle,
and the ex ante perspective. Posner is almost certainly right about the
limited extent to which economics can contribute to the resolution of
ultimate value conflicts. Kaplow and Shavell's contrary claim is far too

---

55. *Id.* at 70-78.

56. *See* Jeremy A. Matz, Note, *We're All Winners: Game Theory, the Adjusted Winner
Procedure, and Property Division at Divorce*, 66 BROOK. L. REV. 1339 (2001).

57. Another solution would be an auction between the family members, but this would
probably be seen as fair only if they could call upon similar resources in the bidding. Also,
bidding could be distorted by the prospect of recapturing some of the bid when the auction
proceeds are distributed to the beneficiaries.

ambitious. But this does not mean that economics is limited to pointing out the range of Pareto-efficient solutions, with nothing at all to say about how to choose between them.

The conclusion, then, is that economics has a limited but valuable role to play in helping to make the legal system more equitable. Economics cannot prove that equity is a valid social goal. But if we do value equity in the legal system, economics may be able to help us define it more carefully and pursue it more effectively. It is a mistake to think that efficiency is the only goal of the legal system that can benefit from economic analysis. Ironically, despite his reputation as an aggressive intellectual promoter of law and economics, this is one area where Judge Posner's claims for economics may be a little too modest. Posner is surely correct, however, that justice cannot be reduced to economics.

Welfarism, as propounded by Kaplow and Shavell, is not a viable alternative. Unless it reduces simply to utilitarianism, welfarism is generally indeterminate. Outcomes will generally depend on the choice of a SWF (and sometimes on whether the function is applied ex post or ex ante). These critical choices — and the constitutional procedures for making them — must be based on norms other than welfarism. And, of course, the underlying individual utility functions are themselves difficult to define. Economics may well contribute importantly to our understanding of how to implement equity. But understanding the ultimate meaning of equity is another matter. For that, we must look outside of economics.

## APPENDIX

For any suitable ranking of outcomes, we wish to construct a corresponding SWF. The technique can be seen most easily in the two-person case, where every utility state assigns one utility to the first person (call her Ann) and another to the second (Mary).[58] Consider any finite set of utility states, ordered according to some preference, $U_N, \ldots U_2, U_1, O$, where O is a point that is Pareto-inferior to all of the others. (For example, if all utilities are positive, O can be the origin.) To ensure that the Pareto principle is respected, we require that none of these states be Pareto superior to a state that is higher in the preference ranking. Now, all we have to do is construct our SWF. The details turn out to be a little complicated, but the process is basically straightforward.

We construct the SWF as follows. Begin with O and assign the following value to any point X which is Pareto-superior to O: $f(X) = 1 - e^{-d(O,X)}$, where $d(O, X)$ is the distance between O and X. Note that $f(O) = 0$, and $f(X)$ rises toward (but never equals) 1 as the distance between X and O increases. (For the rest of the plane, outside of the set of points that are Pareto-superior to O, we use any SWF that is strictly bounded above by 0.) Given any two points that are both Pareto-superior to O, if one is Pareto-superior to the other, then it must be farther away from O, and hence its f-value will be higher. So we have created a SWF on the whole plane that ranks O at the bottom of the listed group, since all the other social states on the list are Pareto-superior to O. Now all we have to do is repeat the process. Recall that our initial SWF was designed to have values less than 1. Now we modify the SWF. We set the f-value of $U_1$ as 1, and for any point X that is Pareto-superior to $U_1$, we set $f(x) = 2 - e^{-d(U1,X)}$. As before, within the set of points Pareto-superior to $U_1$, f constitutes a SWF. Moreover, no point in this new set is Pareto-inferior to any point outside the set, all of which have f-values below 1. So we now have a SWF that ranks $U_1$ above O. We continue the process through each point on the list through $U_N$. We now have a SWF which produces the required ranking.[59] Of course, given a different ranking of points, we could produce another SWF to produce that ranking instead.

These SWFs need some tweaking if we want to add some additional desirable qualities. For example, the SWFs we have defined so far are not continuous at the edges of the Pareto-superior sets for

---

58. The generalization to the n-person case is straightforward but more difficult to visualize.

59. Equivalently, and perhaps more elegantly, we would begin with $U_N$ and work our way back to O, at each step defining the SWF for all points Pareto-superior to $U_M$ but not to $U_{M+1}$. This would avoid the need to first assign and then reassign values to some points.

each of the points $U_N$. To obtain continuity, we need to add a short ramp at the edge of each Pareto set; if we want to get fancier, we can curve the edges of the ramp so that the SWF will be differentiable. In addition, we might also want the SWF to be symmetrical (invariant under permutations of the coordinates). We would then proceed as follows: We take each of the listed utility states and pair it with a state where Ann's utility is at least as high as Mary's, reversing the order of the coordinates where necessary to achieve this. We carry over the original rank ordering to these new points. We must again assume that no point which is higher on the list is Pareto-inferior to a point lower on the list; otherwise no symmetrical SWF is possible. Now, we repeat the same procedure used to create a SWF, as laid out in the proceeding paragraph, except that we define f(X) only if X is above the diagonal line x=y. In the next step, for any point Z below the diagonal line, we match that point with the point X having the same coordinates in reverse order, and set f(Z) = f(X).

If we want to add convexity on top of these qualities, some further complications are needed. To also obtain convexity, we first need a further restriction on the points. For any point U on the list, consider the line between U and its mirror image U', which divides the plane in half. No point lower on the list than U can be in the upper side of this line and no point higher on the list can be on the lower side of the line. Second, we need to include everything Pareto-superior (north or west) of this line in the relevant set when we assign the f-values, rather than simply using the set of points Pareto-superior to U. Third, instead of using d(U, X) in the formula, we need to replace U with V, the point where the line segment between U and U' crosses the x=y diagonal.