

Kaplow, Louis, and Shavell, Steven. *Fairness versus Welfare*.
Cambridge, MA: Harvard University Press, 2002. Pp. 544. \$50.00 (cloth).

Fairness versus Welfare (*FW*) aspires to be the new manifesto for normative law and economics. The authors, Louis Kaplow and Steven Shavell, both of Harvard Law School, are two of the field's leaders. *Fairness versus Welfare* was previously published as an article in the *Harvard Law Review*, after presentation at many faculty workshops and symposia, and has prompted numerous book reviews and responsive articles. Not for quite some time has a work by law professors attracted this much attention. What explains it? Most scholarship in normative law and economics is, in effect, applied ethics—arguing that one or another legal rule is efficient or welfare maximizing. *Fairness versus Welfare* is more fundamental. It seeks to set the agenda for the field, indeed for normative legal scholarship generally. Its central claim: that normative analysis of legal rules (at least by scholars if not governmental officials or laypersons) should focus solely on “social welfare” and should ignore “fairness” considerations. *Fairness versus Welfare* has the same purpose as Richard Posner's famous defense of wealth maximization a generation ago: to defend and entrench a single criterion as the basis for all scholarly evaluation of law.

The position that *FW* defends is, in effect, Paretian consequentialist welfarism plus a full-information preferentialist account of welfare. Legal rules should be evaluated in light of their outcomes. The goodness of an outcome, in turn, is solely a function of individual well-being. Well-being is identified as the satisfaction of fully informed preferences. Kaplow and Shavell assume that preferences are measurable by utility functions, and, if so, their position can be expressed as follows. The goodness of outcome x is determined by a social welfare function $F(U_1(x), U_2(x), \dots, U_n(x))$, where the U_i are individual utility functions measuring each individual's preferences, and where F is an increasing function of the U_i (24). The Pareto principle in its strong variant is therefore preserved. If $U_i(x) > U_i(y)$ for at least one individual i , and there is no individual j such that $U_j(x) < U_j(y)$, then if F is an increasing function of individual utilities it follows that $F(x) > F(y)$.

Crucially, *FW* does not take a position as to the specific form of the social welfare function. Any increasing function of utilities that satisfies an anonymity constraint will do or, at least, is not ruled out by Kaplow and Shavell (24–27). In particular, Kaplow and Shavell are not committed to utilitarianism. *Fairness versus Welfare* is a generic defense of Paretian welfarism, and “fairness” is defined by contrast. A “fairness” approach is anything other than Paretian welfarism: any theory that gives some weight to considerations other than the maximization of some social welfare function (39). This is a big tent. Corrective justice theories of tort law, promissory theories of contract law, participatory accounts of legal procedure, and retributivist views of criminal law all come under the rubric of

fairness, as defined by Kaplow and Shavell, and all are vigorously criticized. Indeed, normative legal scholarship outside law and economics almost always has some nonwelfarist element. So, despite the generic, not necessarily utilitarian, cast of *FW*, its critical target is also very large.

Kaplow and Shavell attack the target with gusto, at times lapsing into a polemical, even scornful tone. And much of the response to *FW* has been unfavorable, often quite hostile. All this sound and fury has obscured the fact that this new brief for normative law and economics is, substantively, a genuine advance over the traditional commitment to wealth maximization—a position that Posner himself has abandoned but that is still reflected in much legal scholarship and policy analysis. Wealth maximization or its close cousin, Kaldor-Hicks efficiency, has little plausibility as a moral criterion, let alone a decisive one. By contrast, the fact that a legal rule increases overall welfare is, pro tanto, a good moral reason to adopt the rule. To be sure, *FW* argues for social welfare as a full framework for evaluation, not merely a pro tanto criterion. But its position, Paretian welfarist consequentialism, is a live position within contemporary moral philosophy and is the dominant view in theoretical welfare economics and social choice theory. Unlike wealth maximization, Kaldor-Hicks efficiency, or utilitarianism, Paretian welfarist consequentialism is potentially sensitive to the distribution of welfare. Again, *FW*'s social welfare function is any function for ranking outcomes that depends solely on individual utilities, increases with them, and satisfies an anonymity constraint. This includes straight utilitarian aggregation, but it also includes a “prioritarian” function that gives greater weight to the welfare of individuals with lower welfare levels.

Fairness versus Welfare departs from traditional law and economics in other ways. Kaplow and Shavell reject or at least decline to defend the following problematic positions, all standardly associated (at least in the past) with law and economics: (1) *Preferences are revealed choices*. Kaplow and Shavell reject this behaviorist view (410 n. 24) and accept the better view that sees preferences as attitudes which may be imperfectly revealed by choices. (2) *The exogeneity of preferences*. Kaplow and Shavell acknowledge that preferences can change and that law can shape preferences (413–18). (3) *The impossibility of interpersonal welfare comparisons*. Skepticism about comparability is what motivated the development of the Kaldor-Hicks standard. Kaplow and Shavell, by contrast, accept that utilities are interpersonally comparable to some extent (24 n. 15). (4) *Antipaternalism*. Since welfare consists in the satisfaction of fully informed rather than actual preferences, Kaplow and Shavell acknowledge that individuals can be mistaken about their own welfare and that government regulation of individual choice might be justified even absent externalities (410–13). (5) *Market solutions*. Kaplow and Shavell do not argue that the market is presumptively better at maximizing social welfare than regulation. The aim of *FW* is to defend a moral view, not the additional claim that this view favors a certain set of institutional structures, for example, market ones. (6) *Distribution through taxes and transfers*. Kaplow and Shavell, in other work, have advanced a particular institutional claim, namely, that distributive considerations are best handled by the tax-and-transfer system rather than regulation, common law doctrines, or other legal rules. In *FW*, Kaplow and Shavell do not reject this claim, but neither does it play a significant role in their argument.

Substantively, then, *FW* represents a pretty plausible view. Deontologists, of course, will rebel at the authors' consequentialism; consequentialists who believe in a consequentialism of rights, or in a responsibility-sensitive consequentialism, will rebel at their strict welfarism; certain egalitarians will take issue with their Paretianism, arguing that equality may require "leveling down"; and rule consequentialists will object to their act centrism. But Paretian welfarist (act) consequentialism should certainly be on the table, and Kaplow and Shavell have cleaved away various economic dogmas, just mentioned, that are inessential to the view.

Why, then, the heated critical response to *FW*? Part of the answer is its polemical tone and the fact that the authors are often uncharitable in their reading of particular fairness theories. *Fairness versus Welfare* includes separate chapters discussing tort law, contract law, legal procedure, and criminal law, which are larded with references to the relevant nonwelfarist scholarship. Still, one feels that Kaplow and Shavell have not sympathetically engaged this scholarship. They haven't really read it. Kaplow and Shavell repeatedly accuse non-welfarists of having an incomplete theory or of failing entirely to address some issue, and these accusations are often overstated. To give but one illustrative example, they complain that corrective justice theorists of tort law have not explained what makes acts wrongful (93–96). This is absurd: although there is as yet no consensus account of wrongfulness in this area, individual corrective justice theorists surely have furnished their own accounts, and although the individual accounts may well be incomplete or unspecified in some ways, so is *FW*'s account of social welfare (which, remember, is any increasing, anonymous function of individual utilities).

Moreover, one of the two main arguments for welfarism that Kaplow and Shavell advance is a nonstarter. The structure of *FW* is the following. An introductory part articulates and clarifies the authors' position, and a concluding section adds further clarification. In the lengthy middle portion of the book, Kaplow and Shavell consider in turn the four areas of tort law, contract law, procedural law, and criminal law. In each of these areas, Kaplow and Shavell argue against fairness in two principal ways. First, they consider salient doctrinal issues, such as the choice between strict-liability, negligence, or no-liability rules of tort law; the choice between specific performance, expectation damages, and reliance damages as the remedy for breach of contract; the choice between procedural doctrines that allow suit, subsidize suit, or prohibit suit; or the choice between a punishment rule that requires the sanction to be proportional to the crime and one that allows disproportionate sanctions. Using a wide range of simple fact patterns, Kaplow and Shavell illustrate how pursuing fairness can reduce social welfare. Second, Kaplow and Shavell argue that the intuitions favoring fairness views are a product of social norms or inborn dispositions. These norms and dispositions have evolved to maximize social welfare and do so both by reducing individual opportunism and by serving as heuristics that ordinary individuals properly employ, given their cognitive limitations and deliberation costs. But the intuitions are no substitute for welfarist deliberation by legal experts, and, indeed—here's the nub of Kaplow and Shavell's second argument—once we understand the origin of these intuitions, their force in justifying fairness views evaporates.

This second argument is a muddle, as Jules Coleman has shown at length (“The Grounds of Welfare,” *Yale Law Journal* 112 [2003]: 1511–43). *Fairness versus Welfare’s* explanation of why norms and dispositions would evolve to maximize social welfare is sketchy. And even if that explanation persuades, Kaplow and Shavell haven’t done the epistemological work to show that explaining intuitions vitiates their role in good moral reasoning. If we’re naturalists about intuitions, they are all explainable, and Kaplow and Shavell’s view would imply that the Rawlsian, reflective-equilibrium account of justified moral beliefs (which has a role for intuitions) is incorrect. Maybe it is, but Kaplow and Shavell haven’t given us a replacement account and, indeed, repeatedly appeal to the reader’s intuitions in favor of welfare and Pareto improvements. Perhaps Kaplow and Shavell might respond, here, that the fairness intuitions they mean to criticize are no more than concrete reactions about particular cases, while welfarist intuitions are systematized through the much more general construct of “welfare.” But fairness intuitions, too, are systematized through general notions like autonomy, responsibility, rights, wrongs, property, promises, and compensation.

What about the first argument for welfarism: that any theory which pursues fairness can reduce social welfare? This may seem to be a tautology. Actually, it isn’t. It isn’t even true. Fairness theories give some role, even a small one, to considerations other than social welfare; thus a tiebreaker theory which gave lexical priority to social welfare over nonwelfarist considerations would be a fairness theory (by Kaplow and Shavell’s definition) and yet never reduce social welfare. So the authors can’t be accused, as some critics have done, of wasting hundreds of pages demonstrating a tautology. But let’s not quibble about the meaning of “tautology.” The leading nonwelfarist accounts of tort law, contract law, adjudicative procedure, and criminal law do not give lexical priority to social welfare over nonwelfarist considerations. Once it is recognized that they don’t, it is immediately obvious that these accounts can reduce social welfare. So are the numerous concrete examples that Kaplow and Shavell deploy in the middle part of their book a waste of time? I think not. The examples are useful in just the way hypothetical cases generally are for moral reasoning, spawning intuitions that might sway nonwelfarists toward welfarism (or vice versa) and forcing nonwelfarists (and welfarists) to further specify their views. For example, what does retributive justice demand where disproportionate penalties lead to fewer crimes and a lower rate of conviction of the innocent (336–39)? Where tort liability doesn’t tend to induce precaution, and the costs of administering a liability system are high, and it seems roughly equally likely ex ante that a given person will be an injurer or a victim, why would corrective justice still demand a regime of strict liability or negligence (100–106)?

Moreover, there is a general claim that Kaplow and Shavell seek to illustrate with their numerous fact-patterns that is far from obvious, namely, that any fairness theory can reduce everyone’s welfare. For any such theory, they claim, there exists some choice situation where the legal rule required by the theory is Pareto inferior to an alternative. In particular, consider symmetric situations, where everyone is identically situated with respect to welfare. In such a situation, the legal rule chosen by a social welfare function that increases in individual utilities will be best for everyone, and any theory which requires a different choice will be worse for everyone (52).

This is a crucial line of argument. The claim that any nonwelfarist theory will, in some possible situation, require a Pareto-inferior choice is an arresting and important claim and constitutes a truly original contribution to the philosophical debate about welfarism. Unfortunately, the claim is underdeveloped in *FW* (52–58). Imagine, once more, a fairness theory that gives lexical priority to social welfare over nonwelfarist considerations. Or imagine the sort of theory that Howard Chang describes, one that generally pursues nonwelfarist considerations but gives lexical priority to the Pareto principle (“A Liberal Theory of Social Welfare,” *Yale Law Journal* 110 [2000]: 173–258). Why would these hybrid theories ever require Pareto-inferior moves? Kaplow and Shavell’s answer, presented most crisply in a technical article, is that the Pareto principle plus a continuity requirement implies welfarism (“Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle,” *Journal of Political Economy* 109 [2001]: 281–86). Continuity means, very roughly, that small changes in some determinant of the social ordering of outcomes or choices don’t produce large changes in the ordering. But why is continuity a morally compelling axiom for the consequentialist, let alone the nonconsequentialist? Consider the “leximin” principle, much mooted in the social choice literature since Rawls: priority in ranking outcomes is given first to the welfare of the worst off, then the second-to-worst off, and so on. A leximin social ordering satisfies Paretian welfarist consequentialism but is discontinuous in individual utilities.

Kaplow and Shavell seem to have concluded that a full-blown discussion of the conditions under which the Pareto principle implies welfarism and a full defense of those conditions as desiderata for any moral theory would have been too difficult for *FW*’s intended audience of law professors. Perhaps so. But, conversely, a book incorporating that discussion and defense would have been of much greater interest to philosophers and social choice theorists.

Two additional observations: (1) Kaplow and Shavell assume that welfare is measurable and interpersonally comparable but say little about how to construct an interpersonally comparable scale of well-being. It is a standard result in social choice theory that, if utilities are interpersonally comparable but represent only welfare levels and differences (rather than, say, ratios), welfarism plus a few other uncontroversial axioms imply that the social ordering is either leximin or a kind of utilitarianism (Salvador Barberà, Peter J. Hammond, and Christian Seidl, eds., *Handbook of Utility Theory* [Dordrecht: Kluwer, 1998], 1:418). So Kaplow and Shavell face the trilemma of requiring a high degree of precision in measuring welfare, abandoning the possibility of a distributively sensitive social ordering, or abandoning their commitment to continuity. (2) Kaplow and Shavell equate welfare with the satisfaction of fully informed preferences, refusing to “restrict” preferences as some preferentialist accounts of well-being do—for example, by not counting the satisfaction of sadistic or otherwise objectionable preferences, or moral preferences, as welfare enhancing (418–43). It is a truism that individuals, even rational and fully informed ones, can choose to sacrifice their own interests. What someone wants and what someone wants for herself are possibly different. Kaplow and Shavell’s position precludes this possibility.

MATTHEW D. ADLER
University of Pennsylvania