Kaplow and Shavell on the Content of Fairness

Richard Craswell*

Louis Kaplow and Steven Shavell have argued forcefully that policy makers should give no weight to what they call "fairness" in choosing legal rules.[1] More precisely, they have advanced two arguments (or sets of arguments): one based on a formal proof concerning the Pareto principle, and one based on a different set of substantive policies. The Pareto argument receives most of the emphasis in their own writings; it is also the one that has drawn the most attention from their critics. But my focus here will be on the less obvious substantive arguments, which—in my view—provide the stronger justification for Kaplow and Shavell's position.

More specifically, Kaplow and Shavell (hereafter K&S) present one of their arguments as resting on a logical proof, to the effect that anyone who accords any weight at all to fairness is thereby committed to supporting at least some rules even under circumstances where everyone in society would be made worse off. As other writers have already pointed out, though, this argument does not work against certain forms of "hybrid" fairness theories: those which (a) assign normative importance to fairness in most cases, but (b) do not do so in evaluating any rule that would make everyone better off. Section 1 briefly explains why K&S's argument does not work against these hybrid theories—but since this object-

---

* William F. Baxter–Visa International Professor of Law, Stanford University. I am grateful for helpful comments and suggestions from Howard F. Chang, Barbara H. Fried, Louis Kaplow, and Steven Shavell.

1. Louis Kaplow & Steven Shavell, Fairness versus Welfare (2002). For convenience, I will follow Kaplow and Shavell in using "fairness" as a single catch-all term to describe all of the moral theories they attack.

ion has already been voiced by others, I will not dwell on it here.

Instead, section II proceeds to highlight an alternative justification for K&S's position, by showing the substantive difficulties faced by any hybrid theory of the sort identified in section I. One of the challenges facing any fairness theorist, of course, is to explain *why* certain actions are believed to be unfair. Thus, K&S's attack on all such theories can be understood as raising the following question of substantive policy: In any case in which a fairness theory would conflict with welfarism, why should we believe that the fairness theory has properly identified those acts that are truly unfair?

I. THE PARETO ARGUMENT

I begin, however, with the argument that K&S emphasize, involving a potential conflict between fairness theories and the Pareto principle. They develop this argument by considering two cases: one in which everybody in society is identical, or is identically affected by the legal rule; and one in which some members of society gain from the rule while other members lose.

A. The Pareto argument in the identical case

The identical case is the easiest for K&S. If everyone in society is identically affected by the rule, then either the rule will make everyone better off or it will make everyone worse off. (A third possibility, involving rules that leave everyone exactly the same, can be set aside as not worth arguing over.) Faced with a choice between these options, a welfarist would of course endorse whichever rule makes everyone better off. In these cases, then, whenever a theory of fairness diverges from welfarism it can only do so by endorsing a rule that leaves everyone worse off. Thus, K&S are correct

that non-welfarists must either occasionally endorse rules that leave everyone worse off, or be reduced to mimicking the recommendations of welfarism, *in the special case of rules that affect everyone in society identically.*

Most rules, however, do not affect everyone in society identically, and Kaplow and Shavell's argument in the non-identical case is more complex. The following section will review K&S's own argument, while succeeding sections point out some difficulties.

B. The Pareto argument in the non-identical case

Whenever some people would gain from a rule but others would lose, the rule cannot be evaluated without at least implicitly balancing the gains and losses to different individuals. In their book, K&S take no position as to exactly how those gains and losses should be balanced. Instead, they follow standard welfare economics and posit that we have a "social welfare function" which reflects whatever tradeoffs society is willing to make in this regard.[2] They emphasize that this social welfare function might well reflect distributive considerations, so that rules whose losses would be borne by poorer members of society might be disfavored for that reason.[3] For purposes of their formal argument, however, K&S are content to leave the exact form of the social welfare function unspecified. Instead, their claim is that legal rules should be selected on the basis of *some* such social welfare function, rather than on the basis of a fairness theory whose recommendations might diverge from those of the social welfare function.

Of course, K&S do not make this argument simply by assuming

2. Id. at 24–27.
3. Id. at 28–38.

that any particular social welfare function is the best way to balance gains and losses to different individuals. If they could start with that assumption, the rest of their argument would be easy (too easy!), for if a particular social welfare function identifies the best distribution of gains and losses, it would then follow automatically that the distribution of gains or losses produced by any other rule must be less than ideal. The reason this simple argument fails, though, is that supporters of a fairness theory could always respond by claiming that they were following a *different* social welfare function, in which the distribution of gains and losses produced by their rule ranked higher than the distribution produced by whichever rule was supported by K&S.[4] If two rules each leave some individuals better off and others worse—in other words, if neither rule Pareto-dominates the other—it will always be possible to construct one social welfare function under which the first rule's distribution of gains and losses rates more highly, and another social welfare function in which the second rule's distribution rates more highly. Thus, as long as K&S are unwilling to commit themselves to defending the exact content of some particular social welfare function, it might seem that their argument would always be vulnerable to this response.

In fact, though, K&S have a more subtle argument, which does not require them to defend any particular social welfare function. Instead, they argue that as long as a social welfare function satisfies a few fairly general properties—properties to which I shall return in succeeding sections—then a social welfare function that also values fairness will inevitably lead to problems. Specifically, K&S

4. Some arguments of this form are made in Michael B. Dorff, Why Welfare Depends on Fairness: A Reply to Kaplow and Shavell, 75 S. Cal. L. Rev. 847 (2002). I discuss an error in part of Dorff's argument in note 13 infra.

claim that any social welfare function that includes fairness will require its supporters to endorse some rules, in some circumstances, that make everyone in society worse off. Even if the particular rule under immediate consideration does *not* make everyone worse off, K&S claim that any fairness-based social welfare function must endorse other rules (under other circumstances) that *do* make everyone worse off.

As this is the key step in K&S's Pareto argument, it is worth examining closely.[5] Suppose we are considering a possible rule— call it *X*—which, let us stipulate, treats some citizens unfairly (according to our best theory of "fairness"). Suppose further that this unfair treatment leads a fairness theorist to oppose rule *X*, in favor of some alternative *Y*. As long as rule *Y* leaves at least some people better off than they would be under the unfair rule *X*, it might seem that the fairness theorist could oppose *X* without any conflict with the Pareto principle.

However, K&S then ask which rule the fairness theorist would endorse if the situation changed slightly, so that rule *X* could now be applied with lower administrative costs, and these administrative savings were just sufficient (and were distributed among society in just such a way) to leave everyone a little bit better off than they would be under rule *Y*. If rule *X* is still an unfair rule—that is, if the savings in administrative costs has not affected the fairness of *X*—K&S argue that the fairness theorist might still have to support rule *Y*, at least in certain cases. True, if the weight given to *X*'s unfairness were small, and if the administrative savings were huge, even a fairness theorist could admit that the unfairness of *X* might

5. This description is based on the argument in Kaplow & Shavell, supra note 1, at 53 n.75. I discuss the more formal, mathematical version of their argument in the text infra at note 15.

in some cases be outweighed by its administrative savings. But as long as fairness receive any weight at all, the fairness theorist would still have to be prepared to oppose *X* in any case where all other considerations left the decision very close to the line—as, for example, if the savings in administrative costs made *X* only just barely superior to *Y* on welfare grounds. Otherwise, if fairness were *never* enough to tip the decision from *X* to *Y,* then fairness would lose any operational significance in the theory.

This, then, is the challenge posed by K&S's Pareto argument. The challenge is not that fairness theorists would *often* be required to support rules that left everyone in society worse off, for such rules are rarely at issue in legal debates. Instead, the challenge comes from K&S's apparent demonstration that fairness theorists would be committed by their beliefs, to supporting some rules that make everyone worse off *if the circumstances leading to that result ever did occur* in actual life. In the view of K&S—and, probably, in the view of a good many others—there is something suspect about a moral premise which implies that it might, even in rare cases, be appropriate to make everyone in society worse off.

C. The Pareto argument and "hybrid" theories of fairness

As other authors have pointed out, though, at least some fairness theories can respond to this challenge.[6] Specifically, K&S's Pareto argument may not work against fairness theories that take the following hybrid form:

---

6. Versions of the argument in this section have been made by Barbara Fried, Can We Really Deduce Welfarism From the Pareto Principle (unpublished manuscript, February 2002); and Howard F. Chang, A Liberal Theory of Social Welfare: Fairness, Utility, and the Pareto Principle, 110 Yale L.J. 173, 208–13 (2000). I will discuss K&S's responses to this argument in the following section.

(1) In most cases, choose which rule to support on the basis of a theory in which fairness does indeed receive weight; but
(2) If there is a rule that would make everyone in society better off, always support that rule.

To be sure, constructing an attractive theory of this form will not always be easy. For one thing, its attractiveness will depend heavily on the substance of whatever fairness theory is embodied in clause (1). In addition, some theories that seem to be grounded in fairness may in fact be based on distributional concerns, which (as K&S point out) are entirely consistent with welfarism. Further, to the extent that clause (1)'s theory of fairness is truly different from welfarism, there may be a risk of intransitive or incomplete rankings, since some rules will be evaluated under clause (1) of the theory while others will be evaluated under clause (2).[7]

For now, though, my point is simply that this hybrid theory is immune to the argument sketched out in the preceding section. That argument confronted fairness theorists with an apparent dilemma by requiring them to support a fair rule *Y* over an unfair rule *X*, even if *Y* would leave everybody in society worse off. Under a hybrid theory, though, this particular dilemma vanishes. If literally *everyone* in society would be better off under *X*, the hybrid theorist can invoke clause (2) of the theory, and support *X* just as any welfarist would. But the hybrid theorist could still continue to give weight to fairness in any case in which the rule being considered would not leave everyone worse off.[8]

7. I will return to some of these technical difficulties infra in note 17.

8. Note that this hybrid theory is not the same as the "mixed" fairness theories discussed by Kaplow & Shavell, supra note 1, at 42–44. In mixed theories, the

Moreover, I suggest that there are a number of plausible (or apparently plausible) fairness theories that are functionally equivalent to the hybrid theory stated above. In particular, consider any theory of the following form:

> (1) In every case, choose which rule to support on the basis of a theory in which "fairness" is indeed entitled to some weight; but
> (2) There is nothing unfair about any rule that leaves everyone in society, including the victims of what might otherwise be unfairness, better off.

Under these theories, fairness is *always* to be taken into account, but there is no unfairness left to consider if the putative "victim" of the unfairness has not been made worse off. As a consequence, these theories will share the "hybrid" feature of giving no weight to fairness if (but only if) everyone is made better off by a rule. And provision (2) of this theory—what we might call the Pareto provision, or the "no harm, no foul" clause—might seem attractive to fairness theorists on any of several grounds.

1. No unfairness if actual consent. For example, consider the class of theories which hold that individuals have certain rights, and that it would normally be unfair to infringe those rights; but which also hold that the rights are waivable, and so can be "infringed" without unfairness if the rights-holder consents. Now consider two possible rules, one that leaves everyone in society better off, and one that does not. In the case of the rule that leaves everyone better off, presumably every member of society would

---

fairness of a rule is never accorded exclusive weight, but it is always given *some* weight, even if everyone would be better off if a different rule were adopted.

consent to that rule (setting aside transaction costs, at least for the moment). As a result, such a rule could be enacted with no unfairness at all—the potential unfairness having been vitiated by everyone's consent—so there would be no unfairness left to consider in this case. By contrast, in the case of the rule that did not benefit everyone, there would be at least one rights-holder who would not consent to that rule, thus leaving some unfairness to be considered. The result is a hybrid theory of precisely the sort that I described above.

2. No unfairness if hypothetical consent. Of course, in a large society it is usually impossible to obtain actual consent from every affected party, if only because of transaction costs. However, the same hybrid characteristic is present in any theory that allows a similar role to be played by hypothetical consent, so any rights-holder who is made better off by a rule can be *presumed* to have consented to what might otherwise be an unfair violation of his or her rights. Under these theories, too, fairness would drop out of the consideration of any rule that did in fact leave everyone better off, while retaining its role in the consideration of any rule that produced both winners and losers.

3. No unfairness without actual injury. Indeed, even without invoking the concept of "hypothetical consent," it would not be unreasonable for a hybrid theorist to believe that unfairness is present only when the victim of the purported unfairness has been injured in some way. For example, some theories of property rights might say that the owner of real property owns that space *usque ad caelum,* implying that the owner's rights would be infringed if so much as a satellite passes over the property. The same theorist, though, could surely say there is no unfairness in such a case as long as the passage of the satellite does not injure the property

owner in any way. Thus, this theorist could give full weight to any unfairness associated with an actual injury to property owners, while giving no weight at all to fairness concerns when evaluating rules that did not result in injury. In the latter case, fairness would drop out of the analysis because no unfairness would be present.

To be sure, this particular route to a hybrid theory ("no unfairness without actual injury") is more complex than it might appear, for the notion of "injury" requires reference to some baseline: injured as compared to what? If someone touches me lightly, without my consent but also without my even noticing the touch, it is easy to say that I have not suffered any injury that is relevant to fairness. But not all theorists will want to say that there is no relevant injury if I am accidentally struck hard enough to cause minor pain, but the same rule that allows this striking also allows me to do various things that I enjoy, which might result in my accidentally striking someone else. That is, even if my overall gain is positive from both activities (those in which I am struck, and those in which I occasionally strike others), some theorists might balk at the idea of summing those gains to say I have not really been "injured" by a rule that permits these activities, or to say that there is no unfairness created by that rule.[9] Still, not *all* theorists would balk at this combining or netting together of different activities.[10]

9. For example, Ernest Weinrib is clearly opposed to this form of aggregation, at least in tort law. Ernest Weinrib, The Special Morality of Tort Law, 34 McGill L.J. 403, 408–09 (1989). See also the discussion in Kaplow & Shavell, supra note 1, at 113–15.

10. For example, George Fletcher has famously argued that these risks *must* be aggregated, at least when each of the actors is acting non-negligently. George P. Fletcher, Fairness and Utility in Tort Law, 85 Harv. L. Rev. 537 (1972). Aggregation of this sort is also unproblematic under most "consent" theories (either actual or hypothetical consent), since an actor deciding whether to consent to any

Those who are willing to do so would, to that extent, have given their theories the hybrid character we are seeking, by eliminating "fairness" as a factor in those cases (but only in those cases) where those who appear to be disfavored by a rule are in fact better off.

4. Related philosophical distinctions. Speaking more generally, many non-welfarist philosophical theories distinguish the moral evaluation that is required in cases in which everybody is made better off, from the evaluation required in those cases where some people lose. In Kantian terms, if I am trying to benefit at another person's expense, I may be treating that person as a means rather than as an end, thus making my actions impermissible in a way that they would not be if my actions made everyone better off.[11] To a similar effect is the famous criticism that utilitarianism suppresses individual integrity, by its demand that actors sacrifice their own goals and commitments (if necessary) in order to serve the greater good of others.[12] The criticism, of course, is not that it is illegitimate to require such sacrifices when they are necessary to advance the welfare of the individual who is making the sacrifices, so even that individual ends up better off. Instead, the objection holds that it is a different thing entirely to require one individual to sacrifice his or her goals and commitments merely to advance the well-being *of others*.

I should emphasize that I am not claiming here that any of these philosophical objections is correct. Instead, my claim (in this section) is merely that it is not at all uncommon, in philosophical

---

rule would normally want to consider *all* of the effects of that rule.

11. A point that is emphasized by Barbara Fried in her response to Kaplow and Shavell. Fried, supra note 6, at 27–30. See also Chang, supra note 6, at 210–12.

12. E.g., Bernard Williams, A Critique of Utilitarianism, in J.J.C. Smart & Bernard Williams, Utilitarianism For and Against 114–17 (1973).

circles, to distinguish between the kind of justification that is required when nobody else is being injured, and the different justification that is required if one person's gain comes at another's expense. As we have seen, a fairness theorist who is willing to invoke that distinction may be able to escape the force of K&S's Pareto argument. As long as fairness plays no part if nobody is made worse off by a rule, the fairness theorist will not be committed (*pace* Kaplow and Shavell) to endorsing rules that make everyone worse off.

Indeed, I suspect that many legal academics who are attracted to fairness theories will soon come (if they have not already done so) to see their own theories as including just such a hybrid component. To be sure, this move will not be available to *all* fairness theorists, for some may not want to support (all) Pareto-superior moves.[13] But many fairness theorists may consider that they give up little of practical importance by conceding an exception for cases in which everybody would be better off, given that such cases rarely arise. And others may even be attracted to such an exception as a matter of principle, for the same reason that the Pareto principle itself is attractive. As noted earlier, there are many philosophical reasons why normative analysis at least *might* be thought

13. Some of the theories considered by Michael Dorff have this characteristic. Dorff, supra note 4, 75 S. Cal. L. Rev. at 881–88.

Though it is not crucial to all of his arguments, Dorff errs when he asserts that these theories would satisfy K&S's definition of "individualistic." For example, Dorff argues that a social welfare function could simply fail to count certain "illegitimate" gains (e.g., gains obtained by violating deontological commandments) without running afoul of K&S's argument. However, this welfare function could violate the Pareto principle by preferring a state of the world in which all individuals had lower actual welfare but no illegitimate gains were present, over a state of the world in which every individual's actual welfare (including his or her illegitimate gains) was higher.

to be different in cases where no citizen is being asked to bear a loss for the good of others. If, in addition, a hybrid component will also shield the theorist from the force of K&S's formal argument, there is no reason for these theorists *not* to adopt a hybrid theory.

D. Kaplow and Shavell's responses

What, then, can K&S say to these newly hybridized fairness theorists? In the initial presentation of their Pareto argument, K&S do not even address hybrid theories, for they frame the question in way that implicitly excludes them. In effect, K&S limit their initial argument to fairness theories in which unfairness is *always* given a negative weight, whether or not any victims of the alleged unfairness are made worse off.

Consider again the argument sketched earlier, involving a choice between rule *X* (which was arguably unfair) and rule *Y* (which was not).[14] The argument was that, if something changes so that rule *X* now leaves everybody just a tiny bit better off than rule *Y*, a fairness theorist would still have to support rule *Y*, as long as the unfairness of rule *X* is to receive any weight at all. This argument is perfectly sound, as long as the unfairness of rule *X* is always given a (non-trivial) negative weight by the fairness theorist. If the weight assigned to *X*'s unfairness is always negative, K&S can posit a case in which the non-fairness advantages of rule *X* are just small enough to be less than that weight, so that the fairness theorist is still compelled to support rule *Y* (the one that leaves everybody worse off than *X*). But K&S cannot posit such a case against hybrid fairness theories, in which the fairness weight falls to zero if none of the "victims" of the unfairness are made worse off.

14. See the text supra following note 5.

This limit on the Pareto argument also applies to the formal mathematical proof that K&S have published separately.[15] There, K&S prove that any nonwelfarist social welfare function must make it possible for some rules to rank higher than others, even though the higher-ranking rules leave everyone in society worse off. However, K&S limit the scope of their proof to welfare functions that evaluate all possible *states of the world,* by assigning some numerical value to every possible state.[16] They therefore do not consider social decision rules that instead evaluate *moves* from one state to another state, thereby excluding any rule that identifies certain *moves* as impermissibly unfair.[17] In particular, they do not consider hybrid theories that follow the Pareto principle to evaluate any move that leave everyone better (or worse) off; but that

15. Louis Kaplow & Steven Shavell, Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle, 109 J. Pol. Econ. 281 (2001).

16. Id. at 283. Using formal notation, they limit their analysis to social welfare functions of the form $W = W(x)$, where x represents any possible state of the world, and the function W is a mapping from states of the world to the real number line.

17. Using the formal notation of note 16, such a rule can be represented as a function $W = W(x, x')$, where the function W evaluates the move from state x' to state x, and is therefore a mapping from any *pair* of states of the world to the real number line. For example, the Kaldor-Hicks principle is based on just such a pairwise comparison of states of the world. (So is the Pareto principle itself.)

To be sure, rules that evaluates *moves* from one state to another will not necessarily generate complete rankings of all possible states, so a hybrid theorist might sometimes be unable to conclude either that state X is superior to state Y, or that state Y is superior to state X. (This incompleteness can also arise, in certain conditions, under either the Kaldor-Hicks principle or the Pareto principle.) As noted earlier, hybrid rules that evaluate moves from one state to another may also fail to be fully transitive. A formal model of a hybrid rule, which is not complete but is transitive in one direction, is available from the author upon request.

follow some other, fairness-based principles to evaluate moves that leave some people better off and others worse.

In short, K&S's main presentation of their Pareto argument simply is not addressed to hybrid theories of the sort that I have considered here. Instead, K&S explicitly consider hybrid theories on only a few brief occasions, mostly in response to criticisms by others. Their longest discussion is this:

> It has been suggested to us that perhaps the conflict between a notion of fairness and the Pareto principle could be avoided if the notion [of fairness] were modified by assuming it to be inapplicable whenever it would conflict with the Pareto principle. (We note that such a modified notion of fairness is not advanced in any legal or philosophical literature of which we are aware and, in any event, would be inconsistent with the spirit of the rationales offered for notions of fairness.) Such a modified notion of fairness would, however, be odd because, under it, trivial changes in facts would alter the entire basis for assessing legal policies. To illustrate, suppose that in regime *Y* everyone is better off than in regime *Z*, but that one person is only very slightly better off, say, by a penny. Now consider a regime *Y′* in which the only difference from *Y* is that this person is not better off but instead is worse off than in *Z* by a penny. Then, under the modified notion of fairness, the normative method of evaluation would change abruptly — from the Pareto principle, based solely on individuals' well-being, to some notion of fairness — as a consequence of a two-cent difference in

one person's position.[18]

I am not sure that K&S are correct when they say that such a "modified" notion of fairness would be inconsistent with the spirit of most fairness theories. True, there are undoubtedly *some* fairness theories whose motivating premises would not allow any exceptions to be made for cases in which all parties end up better off. As I noted earlier, though, and as other authors have emphasized,[19] many philosophical theories do distinguish between the normative method of evaluation that is appropriate when some citizens must be asked to bear losses, as compared to the method of evaluation appropriate to cases where nobody must lose.

In any event, whatever the state of the existing literature, the important point is that K&S are here not challenging the hybrid theories with what I have called the Pareto argument, by trying to prove that supporters of such theories would have to endorse rules that would leave everyone worse off. Instead, in this passage K&S

18. Louis Kaplow & Steven Shavell, The Conflict between Notions of Fairness and the Pareto Principle, 1 Am. L. & Econ. Rev. 63, 72 n.20 (1999).

Kaplow and Shavell also touch briefly on hybrid theories in their response to Howard Chang. In explaining why they limit their focus to what I would call non-hybrid theories, K&S say: "Suppose that *no matter how much the degree of fairness differed between two regimes,* a notion of fairness *never* implied that one regime was superior to another when all else was equal, namely, when everyone had the same level of well-being. Clearly, there is no sense in which the notion of fairness is receiving any independent weight." Louis Kaplow & Steven Shavell, Notions of Fairnes Versus the Pareto Principle: On the Role of Logical Consistency, 110 Yale L.J. 237, 241 n.10 (2000). But this, of course, is exactly the point of a hybrid theory: that fairness does indeed receive no independent weight *in the special case where nobody is made worse off,* while at the same time it continues to receive independent weight *in all other cases.*

19. This point is developed at more length in Fried, supra note 6, at 26–34. See also Chang, supra note 6, 110 Yale L.J. at 210–12.

Kaplow & Shavell on fairness

are challenging hybrid theories on their merits *qua* fairness theories, by questioning the notion of fairness that the theories presuppose. For example, the objection K&S raise here is that hybrid theories have an inevitable "knife edge" character to them, as differences amounting to only a penny or two could take a rule out of the special case in which everyone is better off, and thereby radically change the evaluation of that rule. In effect, K&S are arguing that no notion of "fairness" can be attractive if it allows the evaluation of a rule to change radically on the basis of such a tiny change in underlying circumstances.

Of course, there is room for debate about whether K&S are right in their implicit premise here (i.e., that the fairness of a rule should not radically be altered by tiny changes).[20] Still, the important point for now is that K&S make a number of arguments of this kind: arguments that do not rest on the claim that a fairness theory must violate the Pareto principle, but rather arguments that the content of particular fairness theories are unpersuasive. These non-Pareto or content-based arguments (as I will call them) can easily be overlooked, partly because they are spread throughout different sections of the book dealing with particular areas of law, and partly because K&S themselves give greater emphasis to their Pareto argument. For example, after previewing a range of substantive problems with particular fairness theories, K&S say that "we do not make them [the substantive problems] the focus of our critique because they are not *inherent* in the idea of giving weight to notions of fairness."[21]

20. For a useful (and critical) discussion of this premise, see Fried, supra note 6, at 24–34. See also Chang, supra note 6, 110 Yale L.J. at 222–26, and the discussion in the text infra at note 24.

21. Kaplow & Shavell, supra note 1, at 51 (emphasis added).

There is, however, a risk involved in focusing exclusively on the Pareto argument, or on arguments that seem to appply "inherently" to all theories of fairness whatever their content. As we have already seen, the Pareto argument does not apply to all hybrid fairness theories; and these hybrid theories are likely to be especially attractive to legal academics. The danger, then, is that readers who hold these theories may correctly perceive that K&S's Pareto arguments do not apply to them, and then incorrectly leap to the conclusion that they can disregard the rest of the book as well. As I will argue in section II, that would be a mistake.

II. THE CONTENT-BASED ARGUMENTS

A. Arguments regarding intrinsically evil behavior

One substantive criticism argues that, if an act is intrinsically unfair (or "intrinsically evil," as K&S put it), a fairness theorist cannot say that this intrinsic unfairness somehow disappears in the special case in which no one is made worse off by it.[22] I will not address this argument at any length, for it clearly does not apply to all fairness theories. As we shall see, there are many forms of behavior which might be deemed unfair, under various theories of fairness, even though the behavior is not "intrinsically evil."

B. Arguments regarding discontinuous changes

A more significant substantive criticism is the one we have already

22. Kaplow & Shavell, supra note 1, at 55 n.76 ("if we view some act as intrinsically evil under a notion of fairness, but the notion becomes modified so that we give this evil no weight when such acts make everyone better off (but not when, say, they make almost everyone better off but a single person is ever-so-slightly worse off), a serious question arises whether we can still maintain the view that the act is intrinsically evil.").

met, which points to the knife-edged effect of tiny changes in well-being. If rule *X* leaves any citizen even one penny worse off, it will fail to qualify for the Pareto portion of a hybrid theory, so it will be judged in a way that takes full account of any unfairness to the loser. But if the other effects of the rule improve by only two cents, so that the citizen who was formerly one penny worse off is now one penny better off, *X* will then qualify under the Pareto portion of the hybrid theory, so it will then be evaluated by an entirely different standard (in which "fairness" no longer plays an independent role).

While this objection has some force, it is not likely to persuade every adherent of a hybrid theory. After all, the same objection could be made against nearly any theory in which individual rights play an important role.[23] If you have a right to certain property, for example, most theories of property rights permit me to come within an inch of infringing that right without requiring any permission from you, in which case my actions will be judged under one set of principles—perhaps a negligence standard; perhaps even complete immunity to act however I wish. If, though, I proceed one inch further, I will then be in violation of your rights, which may result in my actions being judged under entirely different principles—perhaps strict liability; perhaps even exposure to injunctive remedies or to criminal sanctions. In other words, well-defined individual rights almost always result in there being a sharp, knife-edged distinction between that which others may do with relative impunity, and that which they may not.

As a consequence, supporters of rights-based theories are not

23. As K&S recognize, libertarian or rights-based theories are included in the "fairness" theories that they mean to criticize. Kaplow & Shavell, supra note 1, at 26 n.18.

likely to be impressed when K&S raise, as an objection, the fact that their theories allow tiny changes in the parties' actions to produce large differences in the evaluative outcome. This is significant, because hybrid theories of the sort that I have described can easily be re-cast in "rights" language. In a hybrid theory, the relevant right would be a right *not to be made worse off* (or, more precisely, not to be made worse off in particular ways that the theorist has identified as unfair). Once the right is defined in these terms, it follows that there is no infringement of any right whenever the victim of an alleged unfairness has not been left worse off. But if the underlying facts change, so the victim is even one penny worse off, the victim's right will indeed have been infringed (under this theory), thus triggering the same knife-edged transition that occurs in almost any theory of rights.

Perhaps for this reason, K&S acknowledge that they do not object to every discontinuity in moral evaluation. For example, they would allow the moral evaluation to change radically depending on whether a promise was broken or not,[24] or (presumably) on whether any other right had been violated. Instead, they assert only that there must be *some* good—perhaps a good completely unrelated to anyone's rights—for which small changes in its distribution cannot produce discontinuous changes in the moral evaluation.[25] But while this assertion seems reasonable at first glance, it overlooks the possibility that (some) rights might be defined so that the holder's overall well-being is an element of any violation, meaning that his rights would not be violated unless he was thereby rendered worse off. Under such a definition, small

24. Kaplow & Shavell, supra note 15, 109 J. Pol. Econ. at 284.

25. Id. at 284–85.

changes in the distribution of *any* good could result in violations of the holder's rights, thus producing a discontinuous change in the resulting moral evaluation.[26]

C. Arguments concerning the content of unfairness

In short, hybrid theorists who see themselves as protecting individual rights are unlikely to be troubled by K&S's "knife-edge" objection. However, these theorists should indeed be troubled by other substantive objections that K&S raise. As is well known, any theory that rests on individual rights must (if it is to be complete) find a way to justify the assertion that individuals do, in fact, have whatever right is at issue. It is one thing to say that "*if* individual A has a right not to be made worse off in the following regard, then any rule that leaves her worse off should be regarded as an unfair infringement of *A*'s rights, and should be evaluated in a way that takes this unfairness into account." But such a claim always implicates a logically prior question: "Why should we believe that *A* has a right not to be made worse off in this particular respect?" And for fairness theories that are not framed in terms of rights, the analogous question is, "Why should we believe that the acts you identify as unfair are, in fact, unfair?"

In my view, these are the most serious criticisms that K&S raise against hybrid fairness theories. Rather than discussing these criticisms in the abstract, however, I will illustrate them with three

26. In their response to Howard Chang, K&S also argue that their continuity assumption requires only "that the weight given to a notion of fairness is not infinitesimal." Kaplow & Shavell, supra note 18, 110 Yale L.J. at 242 (2000). In a hybrid theory, however, the "weight" given to unfairness can indeed become infinitesimal—or, equivalently, the unfairness itself can disappear—but only in the special case in which nobody is made worse off by a rule.

examples. Following K&S's own organization, one example comes from contract law, one from trial procedure, and one from torts.

1. The content of fairness theories in contract

Consider a rule which allows some parties to break their contracts while paying less than the full expectation interest of the other party to the contract—for example, the rule of *Hadley v. Baxendale*.[27] Now imagine a theory of fairness which holds it to be unfair for any breaching party not to compensate the other party for the full value of their expectation interest. If this theory is actually a hybrid theory—that is, if it is applied only in cases where some people are made worse off by the rule—what objection could K&S raise?

In some cases, of course, K&S might not have any objection at all. If the theory is a hybrid theory, this means that fairness would receive no weight at all (and the resulting analysis would be no different from K&S's welfare analysis) if it turned out that the rule left everybody better off. For example, if victims who failed to receive full compensation were nevertheless benefitted on balance, by paying a lower price when they entered into the contract, the rule might well leave everybody better off, in which case it would fall on the Pareto side of any hybrid theory.

Indeed, in the simplest contractual settings, legal rules can easily leave everybody better off—or, if the rule is a bad one, leave everybody *worse* off—because of the potential for adjustments in the contract price. If, for instance, a rule initially benefits sellers, sellers should then be willing to enter contracts at a lower price. Moreover, if the gains to sellers exceed the loss to buyers, the price should stabilize at an intermediate level which divides the benefits

27. 156 Eng. Rep. 145 (1854).

between the parties, and thus leaves sellers *and* buyers better off. In these simple settings, when the contract price is freely adjustable, many contract rules will fall under the Pareto portion of any hybrid theory, so there will be no difference between the analyses of hybrid theorists and pure welfarists.[28]

Still, it is easy to construct more complicated settings in which it will no longer be true that everyone will be better off (or everyone will be worse off). In general, if different contracting parties differ in their preferences, but if the law for some reason must have the same rule for all parties, then whichever rule the law picks will inevitably help some parties while hurting others. As these are the cases in which hybrid fairness theories diverge from pure welfarism, because the Pareto portion of a hybrid theory will no longer apply, these are the cases I will focus on here.

Suppose, then, that it is deemed unfair if the law fails to fully compensate the expectation interest of any victim of a breach of contract. But suppose now that some victims would suffer large losses from breach, while others would suffer much smaller losses. For example, suppose that the rule in question concerns sellers' liability for consequential damages from a breach of warranty, and suppose that some buyers would lose huge amounts if the product fails while other buyers would lose only a tiny amount. Perhaps the high-risk buyers are using the product as a crucial component of an extremely valuable business, while the low-risk buyers are using the same product in ways that are less crucial (or in businesses that are less profitable). Similar differences could also arise in consumer contexts—for example, consumers who travel

28. For further discussion of this point, see Kaplow & Shavell, supra note 1, at 201–03; or Richard Craswell, Passing On the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships, 43 Stan. L. Rev. 361 (1991).

with lots of expensive clothes and luggage will suffer large losses if airlines destroy their baggage, while consumers who travel with fewer or less expensive goods will lose relatively little.

Finally, suppose that these differences in the potential losses cannot be perfectly adjusted for by differences in the price that sellers charge for their goods or services—either because sellers do not know (and could not easily find out) how much each buyer has at stake, or because it's simply too cumbersome for sellers to adjust their prices on a case-by-case basis.[29] In that case, if sellers are held fully liable for *all* damages (however great), they will respond by raising their prices by an *average* amount, to cover their potential liability to their entire run of customers. As a result, the high-risk buyers (those with large potential losses) will benefit from the rule, for they will end up with full insurance for their losses while paying only an average "premium" in the form of a higher price. At the other end of the scale, however, the low-risk buyers could end up worse off from this rule, for they will have to pay the same average price for an "insurance policy" that, for these buyers, has a smaller potential payout.[30]

How would a welfarist and a hybrid theorist assess this rule? The welfarist, of course, would look at the gains and losses to all of

29. Similar effects would arise if there are legal restrictions on charging differential prices, either directly through traditional price regulation, or indirectly through anti-discrimination rules. These effects are usefully analyzed in Christine Jolls, Accommodation Mandates, 53 Stan. L. Rev. 223 (2000).

30. For more detailed analyses of similar examples see, e.g., Gwyn D. Quillen, Contract Damages and Cross-Subsidization, 61 S. Cal. L. Rev. 1125 (1988); Ian Ayres & Robert Gertner, Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules, 99 Yale L.J. 87, 108–12 (1989); Lucian Arye Bebchuk & Steven Shavell, Information and the Scope of Liability for Breach of Contract: The Rule of *Hadley v. Baxendale,* 7 J. L., Econ., & Org. 284 (1991).

the different groups of customers. She would also consider the identity of the different groups, including any other claim each group might have to greater weight in her social welfare function. For example, if the victims with low potential losses (who might be made worse off by the rule) tended to be relatively poor,[31] that might tip her evaluation against this particular rule. Obviously, though, the welfarist's ultimate judgment would depend on how her particular social welfare function balanced the gains and losses to all of the affected parties.[32]

Now suppose that the hybrid fairness theorist evaluates the rule in some way that leads him to a different conclusion from that of the welfarist.[33] (If the two theorists were to arrive at the same conclusion, there would be nothing left to argue about, so the interesting case is the one in which they differ.) To make the example concrete, suppose that the welfarist would on balance favor a lower measure of damages, perhaps because she values reducing the price to the low-risk customer. At the same time, suppose that the hybrid theorist favors a higher measure of damages, perhaps because he puts some independent weight on the unfairness of the law's failure to fully compensate victims of breach.

As we have already seen, such a hybrid theorist cannot be crit-

31. This will often be the case when consequential damages are at issue, for smaller businesses and relatively poor consumers tend (broadly speaking) to have less to lose than do larger businesses and wealthier consumers. For an argument against consequential damages for lost income, based explicitly on this distributive effect, see Richard L. Abel, A Critique of American Tort Law, 8 Brit. J. L. & Soc'y 199, 202–06 (1981).

32. For K&S's own analysis of a similar example, see Kaplow & Shavell, supra note 1, at 213–15.

33. For convenience in the use of pronouns, I assume that my hypothetical hybrid theorist is a male while my hypothetical welfarist is a female.

icized on the ground that his position logically requires him to support any rules that would make everyone worse off. As long as he is truly a hybrid theorist, the Pareto portion of his theory allows him to avoid this criticism. Nor can the hybrid theorist necessarily be criticized on the ground that his theory leads to a discontinuous difference in outcomes, if the high-risk buyers are made even a penny worse off by the rule. If the hybrid theorist seriously believes that the high-risk buyers have a moral *right* to full compensation—or, more precisely, a right not to be left worse off by any rule that denies them full compensation—then he will not be troubled by a knife-edged result that turns so critically on whether their moral rights have been violated.

The obvious question, though, is why such a theorist would belive that high-risk buyers do have a moral right to full compensation. Or, if the hybrid theorist does not ground his stand in a theory of rights, why does he believe that it is *unfair* to deny the high-risk buyers full compensation?[34] Specifically, why is it unfair to deny them full compensation in this particular setting, where giving them full compensation means that the other, lower-risk customers will have to pay higher prices as a result? In settings where the price cannot be individually adjusted, can the high-risk customers really claim a moral entitlement to have the insurance for their potential losses be subsidized by the other, lower-risk customers?

To be sure, there are moral arguments that could conceivably be made on behalf of this claim. As K&S recognize, the hybrid the-

34. For convenience, I will use the term "entitlement" (as in "an entitlement to full compensation") to encompass both the rights-based and non-rights-based versions of a fairness theory. In the remaining discussion, nothing of substance will turn on this distinction, so I will usually treat both versions together.

orist might claim that failing to provide full compensation to the high risk victims (or, indeed, to any victims) would be an affirmative wrong on the part of the legal system. By contrast, the hybrid theorist might argue that any undesirable consequences, such as higher prices to the low-risk buyers, would be merely that: undesirable consequences, to be sure, but not an affirmative wrong committed by the legal system, and hence not a violation of anyone's moral entitlements. In effect, the theorist might argue that these undesirable consequences result not from the legal system but from the price system, and (in particular) from the imperfection in the price system that prevents sellers from adjusting their prices individually for different classes of buyers.[35] While that consequence would be regrettable, the fairness theorist might say, it cannot justify the legal system in affirmatively violating the entitlements of high-risk buyers by failing to pay them full compensation for their losses. In terms of the philosophy literature, this theorist would be invoking the distinction between actively causing harm and merely failing to prevent harm, or the related distinction between agent-neutral and agent-relative duties.[36]

However, these philosophical distinctions are not *necessarily* enough to rescue the hybrid theorist. While both distinctions have

35. This possible argument is considered briefly in Kaplow & Shavell, supra note 1, at 47 n.63.

36. On the distinction between affirmatively causing harm and merely failing to prevent harm, see, e.g., Judith Jarvis Thompson, Killing, Letting Die, and the Trolley Problem, in William Parent (ed.), Rights, Restitution and Risk: Essays in Moral Theory (1986); or Warren S. Quinn, Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing, 98 Phil. Rev. 287 (1989). On the distinction between agent-neutral and agent-relative duties, see Samuel Scheffler, The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions (1982).

often been discussed in the philosophy literature, both are controversial—for example, it is notoriously hard to define the line between "causing" harm and "failing to prevent" harm—even as used to evalute the morality of individual actions. Still further difficulties arise if we try to use those concepts to evaluate the morality of government actions, such as the choice of legal rules. After all, any injuries caused by breach of contract can always be said to be caused (in the first instance) by the party who commits the breach, not by the government—but in that case, no matter which rule the government chooses, it will never be guilty of anything more than merely failing to prevent a harm that is actually caused by others. In this context, then, the distinction between actively causing harm and merely failing to prevent it provides very little support for this particular theory of fairness.

More fundamentally, these arguments still presuppose that the failure to provide full compensation for breach of contract would indeed violate a moral right of the high-risk customers—or, in non-rights terms, that the failure to provide full compensation would be *unfair*—for only on that supposition could it possibly be said that the government had participated in any wrong. But this supposition—that high-risk customers truly have an entitlement to full compensation, even in circumstances in which other buyers would have to pay more as a result of that entitlement—is precisely what K&S's substantive critique calls into question. In other words, we are still lacking any affirmative argument as to why moral theorists ought to recognize *any* sort of right to full compensation, at least in this particular setting.

In any event, K&S also identify another difficulty with any argument for recognizing such an entitlement for high-risk customers. As K&S emphasize, any rights arising out of a contract are

Kaplow & Shavell on fairness

usually thought to arise from the contract itself.[37] After all, without the contract a customer would (in most cases) have no right to any goods or services at all, much less a right to any particular remedy if the goods or services were defective. In most cases, though, the contract will not have addressed the question of remedies either way, so it will be at best an open question as to whether the contract really does create a right to any particular remedy. True, it is always possible to read a contract that is silent as to remedies as creating in the buyer a right to full compensation in the event that the product is defective. But if the contract is truly silent on the question, then it is equally possible to read it as creating in the buyer a more limited right. In other words, any decision to read the contract in one way, as creating one set of entitlements, rather than reading it as creating a different and more limited set of entitlements, surely requires an affirmative justification. And that justification cannot rest on the premise that "the buyer is *entitled* to full compensation," because the existence or nonexistence of that entitlement is precisely what K&S have called into question.

To summarize, rights-based theories always require some justification for the particular right that they assert; just as fairness theories always require some justification for identifying particular actions as unfair. However, such a justification is likely to be particularly difficult if the claimed entitlement is one that (if it exists at all) was created by some joint enterprise, such as a contract, which (a) had the power to bring into existence any number of possible entitlements, but (b) was not explicit about which entitlement it actually chose. This difficulty is perhaps most appar-

37. Kaplow & Shavell, supra note 1, at 163–64.

ent in contract law, where it has been extensively discussed in the relevant literature.[38] As the following example will illustrate, however, similar difficulties arise in many other fields.

2. The content of fairness theories in procedure

Consider now a theory of procedural fairness. To make the example concrete, let us stipulate that it would be unfair not to allow a plaintiff to introduce all relevant evidence that might bear on his or her right to recover damages. Now imagine a proposed legal rule that would, instead, limit the evidence that a party could introduce by excluding some admittedly relevant evidence.

As K&S discuss, there are circumstances under which a welfarist would support such a rule.[39] To be sure, the welfare analysis is complex, for reducing the accuracy of individual liability findings has potential costs as well as potential benefits. On the cost side, less accurate findings will clearly be costly to particular plaintiffs who otherwise would have been entitled to recover, but whose true losses will go uncompensated under the new rule. In addition, less accurate findings of liability can also undesirably reduce the law's deterrent effects on potential defendants. But, depending on the exact circumstances, these costs might be small compared to the costs that would be saved at trial by not having to consider every conceivable piece of evidence in an unrestricted evidentiary hearing. And if the costs of trial were reduced by enough, that might lead to other benefits—for example, the benefit of allowing more plaintiffs to sue (if there were any deserving plaintiffs who

38. For my own contribution to this literature, see Richard Craswell, Contract Law, Default Rules, and the Philosophy of Promising, 88 Mich. L. Rev. 489 (1989).

39. Kaplow & Shavell, supra note 1, at 249–58.

might otherwise be barred from suit entirely by the high cost of trials), or the benefit of an increased deterrent effect on defendants if the probability of being sued were to rise.

In short, while there is no reason that a welfarist would *have* to favor such a rule, suppose that in this case the balance of welfare considerations does tip in favor of the rule. At the same time, suppose that even a hybrid fairness theorist would be led to oppose such a rule, based on the independent weight he would assign to the unfairness of barring relevant evidence. Even a hybrid theorist would have to take this unfairness into account, as long as some people gained from the rule while others were hurt by it, so that the rule did not fall on the Pareto side of the hybrid theory. (In this case, there presumably would be some plaintiffs who were hurt by the rule and other plaintiffs who gained—e.g., those who might not otherwise have brought suit at all—and similarly with potential defendants, who might also gain or lose.) As a result, even a hybrid theorist would have to give independent weight to the unfairness associated with barring relevant evidence. To give us a conflict between the two theories, let us suppose that the weight given to unfairness is large (or the net welfare benefits from the rule are small), so a fairness theorist would oppose this rule while a welfarist would support it.

Once again, K&S's Pareto argument will not dissuade the hybrid theorist here, because the hybrid nature of his theory means that he would not oppose the rule if it turned out to make everyone better off (for example, if the system-wide savings from the rule were so great that even the disappointed plaintiffs benefitted in the long run). Nor will the hybrid theorist necessarily be troubled by the "knife-edge" problem that K&S point to, in which the evaluation of the rule could change drastically based on a

small change in the underlying facts. After all, if individuals really do have a moral right to present all relevant evidence—or, more precisely, a moral right not to be made worse off by being barred from presenting such evidence—then it will of course be crucial to the hybrid theorist whether those individuals have in fact be made worse off, for only in that case will any rights have been infringed. Thus, once we allow the hybrid theorist to define the relevant entitlements in terms of whether anyone has been made worse off, the relevance of the knife-edged line between "worse off" and "not worse off" will follow as a matter of course.

Instead, the more serious question that K&S raise is the logically prior one, which asks why we should recognize such an entitlement in the first place. In particular, why do we recognize such an entitlement under these precise conditions, where (we have assumed) recognizing that entitlement means that other deserving plaintiffs will be barred from bringing suit at all, and there may also be more violations committed because the law's overall deterrent effect will be weakened? Questions such as these do not establish that this entitlement necessarily should *not* be recognized—but they do remind us that some affirmative justification of this entitlement is needed.

From here, many of the possible arguments are similar to those discussed in the contracts example. For example, the hybrid theorist might respond by claiming that the undesirable consequences (some plaintiffs barred from bringing suit at all) are ones that his rule merely fails to prevent, as opposed to ones that his rule plays an affirmative role in bringing about. As in the earlier example, though, this response relies on a distinction that is always problematic, especially for government actions. If the government adopts a rule that makes the costs of legal procedings

high, thereby blocking some plaintiffs from suing, is the government merely "failing to prevent" injury to those plaintiffs? If so, why isn't the government also merely "failing to prevent" an injury when it adopts rules that allow other plaintiffs to sue, but blocks them from introducing the evidence they need to prevail?

More generally, even if the undesirable consequences can be dismissed as a reason *not* to recognize an entitlement to introduce all relevant evidence, we do not yet have any affirmative reason *in favor* of recognizing that entitlement. There are few natural rights theories that would speak to this issue, at least at the required level of detail, for "the right to introduce specific kinds of evidence in a trial" hardly seems on a par with life, liberty, and the pursuit of happiness; or with any other natural rights that might be thought unproblematic.[40] It may be that a justification could be found in the relevant substantive area of law, so if we are designing rules for torts lawsuits (for example), perhaps a theory of tort law would explain why individuals are entitled to use all relevant evidence.[41] Clearly, though, some such justification will have to be supplied.

In many respects, the problems here are similar to those already discussed in the contracts example. In the contracts example, one difficulty was that any entitlement that might exist had to be created by the contract itself, and this led us to ask why the contract should necessarily be read as creating one entitlement rather than some other, more limited entitlement. Here, if we reject a "natural rights" theory of procedural fairness, then any en-

40. As K&S point out, theories of fairness that might bear on procedural issues are generally less well-developed than those in other areas of law. Kaplow & Shavell, supra note 1, at 228 n.6.

41. A point also recognized by Kaplow & Shavell, id. at 253 n.67. I consider a torts example in the following subsection.

titlement we might recognize must be created by the procedural system itself. The analogous question, then, is why the procedural system should be interpreted as creating one particular entitlement (the one invoked by the fairness theorist) rather than as creating some other, more limited entitlement.

More specifically, the question in each example is whether we should recognize one particular entitlement *given the existence of other limits or imperfections* in the system, which means that the protection of this particular entitlement will impose costs on others. In the contract example, the inability of sellers to tailor their prices to each individual customer means that a right to full compensation for high-risk customers would raise the price that low-risk customers had to pay. In the procedural example our inability to control or subsidize the costs of suit means that a right to introduce all relevant evidence (in those suits that do get brought) could raise the cost of legal proceedings to other parties, and possibly foreclose some suits entirely. In each case, then, the recognition of a certain right in one party causes ripple effects throughout the entire system, and some of those effects would be unappealing.

Interestingly, almost every theory of rights already recognizes this problem for effects that are felt entirely in two-party interactions. It is always recognized, for example, that your right to freedom of action may occasionally conflict with my right to physical safety; or that your right to the use and enjoyment of your land may occasionally conflict with my right to the use and enjoyment of mine. As a consequence, almost any theory about what content to give those rights has to somehow come to grips with the need to balance the immediately opposing interests.

Viewed from this perspective, what K&S have done is to emphasize that there are usually far more interests at stake than the

two that are most immediately opposed, and that these other interests may have an equal claim to consideration. That is, whenever we are designing legal rules, there will usually be effects not only on the two parties to the immediate lawsuit (or the two parties to the immediate contract), but also on other litigating or contracting parties. A welfarist would, of course, consider all possible effects on all affected parties. Indeed, a hybrid fairness theorist might consider all of these effects too; but he would also want to consider an additional factor—the unfairness of the rule—whenever someone has been made worse off in a way that violates one of their entitlements. My claim here is simply that this latter move is unavailable to the fairness theorist at the logically prior stage, which K&S should also be read as challenging. That is, whenever we are trying to decide what entitlements ought to be recognized in the first place, we cannot use the assertion that someone's entitlements have been violated as a reason for answering that question in one way rather than another.

3. The content of fairness theories in tort

As a final example, consider a fairness theory that might be applied to tort law. Specifically, suppose that it is deemed unfair for tort victims to suffer bodily injury without receiving full compensation for their losses. Under this theory, any rule that limits victims' ability to receive compensation will have a strike against it from a fairness standpoint.

Now, from a welfarist standpoint, limits on victims' right to compensation might (in some circumstances) be desirable. For example, limits on victims' right to recover could conceivably increase victims' incentives to take their own precautions against accidents, or they might reduce litigation costs in ways that would

yield long-run benefits. Suppose, to make the conflict stark, that the net welfare effects of some particular limit on victims' right to recover just barely favor that limit. But suppose also that these net welfare benefits are fairly small, so any theorist who placed independent weight on the unfairness of limiting victims' rights would have to oppose that rule. If so, then a hybrid theorist would also have to oppose the rule, provided only that the welfare benefits of the rule were not distributed in such a way as to leave everyone better off. In short, let us stipulate that this rule would be supported by welfarists, but opposed by hybrid fairness theorists.

Most of the arguments that might bear on this disagreement have by now been discussed, so we can dispense with them quickly. Clearly, K&S cannot argue that any hybrid theorist who supports this rule would have to support other rules that make everyone worse off, for hybrid theorists are immune to that objection. Instead, K&S's stronger argument will be a substantive fairness argument, in which they question whether this particular limit on victims' right to recover ought to be regarded as unfair. Yes (they might concede), such a rule would be unfair *if* we grant the premise that victims always have a moral entitlement to full recovery for bodily injury, in every possible circumstance. But they would go on to ask the obvious question: Why should we recognize such a moral entitlement in the first place? In particular, why should we recognize such an entitlement if doing so would lead to all of the bad consequences that (we have assumed) would be present in this particular setting?

Now, it might seem that the fairness theorist has a better reply here than he had in either the contract or the procedure examples. In contracts and procedure, it is hard (though perhaps not impossible) to find natural rights arguments as to why some legal in-

stitution—a contract, or a set of rules for conducting trials—should necessarily be designed to bestow one set of entitlements rather than some other, more limited set. In torts, though, it might be thought that there is a more plausible argument for a natural or pre-political "right to bodily integrity," the existence of which does not depend on its having been created by any legal institution. If so, then the fairness theorist can respond to K&S with a substantive theory of tort law—to wit, that it is unfair not to compensate tort victims for their bodily injuries because that would violate the victims' right to bodily integrity.

However, even this response does not completely avoid the substantive argument of K&S. Even if we stipulate that people have a "right to bodily integrity," any right that is expressed in such general terms will have to be fleshed out in a good many details before it can be used to answer specific questions of legal policy, such as whether or when it might be unfair to limit victims' compensation. For example, few people would argue that a "right to bodily integrity" implies that it would be unfair not to compensate me if I have deliberately flung my body into the path of an oncoming car, leaving the driver no chance to avoid hitting me. As a result, whenever we get around to defining the exact contours of the "right to bodily integrity" (and any attendant right to compensation), we will usually have to balance my claims to bodily integrity against others' legitimate claims to potentially conflicting rights, such as their right to drive automobiles in a non-negligent manner.

This means that all of K&S's welfarist objections can be reintroduced, when we are trying to specify just what the "right to bodily integrity" does and does not include, or (equivalently) when we are trying to decide exactly when it would be unfair to compensate

anyone whose bodily integrity has been infringed.[42] Is it unfair to deny me compensation if I have deliberately thrown myself in front of an oncoming car? Presumably not. In that case, is it necessarily unfair to deny me recovery if allowing full recovery would drive up the cost of trials, thereby preventing some other victims from recovering? Surely this is at least arguable. And this is the key point: none of these arguments can be settled simply by asserting that "people have a right to recover for bodily injury," because assertions of a *general* right to recovery tell us nothing about the exact circumstances to which that general right should extend.

In summary, the problem here is ultimately the same as it was in the contracts and procedure examples. Any theory of fairness requires a substantive justification for viewing certain acts (but not others) as unfair. In deciding what should count as unfair, most fairness theorists are perfectly willing to consider those tensions or potential conflicts that are immediately apparent, such as the tension between your right to swing your fist and my right not to be punched in the nose. But once this is granted, it is hard to see why fairness theories should not be equally concerned with more indirect or system-wide tensions, such as the tension between the right to recover full damages at trial and the risk that this might make trial too expensive for many plaintiffs, or between the right to recover full damages and the higher prices that customers may have to pay in future. Is it merely that our moral intuitions are more used to dealing with simple, direct conflicts (your fist versus my nose, or your use of your property versus my use of mine); and that most of us do not have ready intuitions about more complex

42. Note that if any "natural rights" argument were to be advanced in contracts or procedure, a similar difficulty would probably arise there as well.

tradeoffs, such as the effect of cross-subsidization across different groups of customers? If that is the explanation, then surely K&S are right when they urge that simple moral intuitions ought always to be reconsidered if they conflict with more reasoned analyses.[43]

On the other hand, if a fairness theorist *is* willing to consider all of these remote or system-wide tradeoffs, his analysis will inevitably end up looking a lot like the full-blown welfare analysis of K&S. Indeed, the fairness theorist will not at this stage be able to assign any independent weight to the fairness or unfairness of a rule, for at this stage of the analysis we are still trying to figure out what acts or rules should *count* as unfair. As a result, since unfairness cannot play any independent role at this stage, it is hard to see why a distinctively "fairness" analysis would ever differ from welfarism or consequentialism broadly defined.

Of course, this is not to say that there are not plenty of interesting disagreements still to be had, even within the broad rubric of consequentialism. For example, once the effects of various rules have been identified, there is still room for argument about how those effects should be valued. (Is a 1% reduction in deaths from auto accidents worth a $5,000 increase in the price of cars? What about a $20,000 increase?) Kaplow and Shavell are, in general, committed to valuing these different effects just as the individuals who feel the effects would value them, at least to the extent of re-

43. Kaplow & Shavell, supra note 1, at 62–81.

This is as good a place as any to note that some "fairness" theorists are not trying to defend a full-blown normative theory of fairness, but are merely trying to present a coherent interpretation of the best theory that might be said to underly existing legal doctrine (which would not rule out the possibility that existing doctrine might itself rest on incomplete moral intuitions that are not entirely defensible). As K&S are not concerned with this use of fairness theory—see id. at 91–92 and n.12—I will not address it here.

garding it as undesirable to impose on any individual an outcome that the indvidual does not value highly, if some other outcome could instead be provided that the individual would value more highly.[44] But they note that most of their analysis would still be relevant even for many theorists who were committed to some other method of valuation—for example, theorists who want to maximize human satisfaction or the good life as defined *by the theorist's own set of values,* regardless of whether those values were shared by the individuals who would actually be affected.[45] That is, even this theorist should choose legal rules *only* on the basis of their impact on human welfare (now meaning welfare as this particular theorist defines it), without giving independent weight to any non-welfaristic considerations.

III. CONCLUSION

In this essay, I have not tried to canvass every argument that K&S make in their book. To mention just one of the omissions, K&S devote a good deal of space to showing how certain widely-held intuitions about fairness might have arisen as heuristics or rules of

44. Cf. Kaplow & Shavell, supra note 1, at 18–19 ("The only limit on what is included in well-being is to be found in the minds of individuals themselves, not in the minds of analysts.") (footnote omitted). See also id. at 409–13, discussing the appropriate analysis if individuals are misinformed about how different outcomes might affect their well-being.

45. Id. at 23 n.14; see also id. at 87–88 n. 5 (discussing just such a theory in Gregory C. Keating, Reasonableness and Rationality in Negligence Theory, 48 Stan. L. Rev. 311 (1996)). Kaplow and Shavell criticize Keating on the grounds that his proposed analysis would apply even in settings where all individuals were left worse off as a result. It is less clear whether K&S would also object to a hybrid version of Keating's theory, in which the theorist's own values would govern only when doing so helped some individuals while hurting others.

thumb, which produce desirable welfarist consequences in many situations, but which ought to be overridden when they do not.[46] This is an interesting (and no doubt controversial) argument, but it is not one I address here.

Indeed, I also do not intend to entirely dismiss the Pareto argument of K&S. There are, after all, some settings where (to a first approximation) all individuals may well be affected identically by a rule—for example, symmetric accident cases where any individual is equally likely to be an injurer or a victim; or symmetric contract cases where the same individuals are found as buyers or sellers; or (for that matter) almost any contract case in which preferences are homogeneous and prices can freely adjust to reflect the legal rule. At a minimum, settings such as these raise interesting hypotheticals for moral theorists, who would do well to consider whether their theories might leave everyone worse off.

In most current legal debates, however, none of the rules in question would leave everyone better off. Moreover, most moral theorists will find it easy to avoid hypotheticals in which everyone would be made worse off, simply by moving to a hybrid form of their preferred moral theory. Hybrid theories are not necessarily subject to K&S's Pareto argument, so theorists who make this move might believe that they have thereby escaped all of K&S's criticisms.

As I have tried to show in this essay, that belief would be mistaken. In addition to their more prominent Pareto argument, K&S also make a number of substantive criticisms aimed at the content of particular fairness theories. All of these criticisms point to various undesirable consequences that a particular theory might lead to: consequences that a welfarist would clearly want to take into

46. Kaplow & Shavell, supra note 1, at 62–81.

account. But all of these criticisms can also be seen as going to the substance of the fairness theory itself, by asking why the act or the rule in question should in that case be regarded as unfair.

In my view, these substantive criticisms pose the most serious challenges raised by K&S. Some fairness theories ground the charge of unfairness in the violation of rights said to be infringed by the unfair acts, but any rights-based theory must explain why those particular rights (rather than some other, more limited set of rights) are the ones the legal system should recognize. Moreover, it is generally recognized that defining and selecting the rights to recognize will usually require some accomodation of competing interests of others, as when one person's use of her property detracts from another's use of his, or one person's free movement of her own body inflicts a risk of injury on another's. Viewed in these terms, K&S's substantive criticisms can be seen as identifying a set of additional accomodations that might equally deserve consideration, as when compensating one person's injuries might raise prices for everyone else, or when adopting one measure of damages might lead to reduced deterrence and to a greater number of accidents that injure others. At least at first glance, these tensions and tradeoffs seem no less worthy of consideration, in any fairness or rights-based theory, than do the more immediate tensions that are already considered by those theories.

Obviously, I cannot claim that no fairness theorist could ever come up with a satisfactory response to K&S on this point. What I can claim, however, is that some such response is required. In that respect, K&S have made an important contribution to legal debate —and a contribution that can stand entirely independently of the argument based on the Pareto principle.