ISSN 1045-6333

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

NOTES ON WELFARIST VERSUS DEONTOLOGICAL PRINCIPLES

Louis Kaplow
Steven Shavell

Discussion Paper No. 460

02/2004

Harvard Law School
Cambridge, MA 02138

JEL Classes D63, H43, K00

Notes on Welfarist Versus Deontological Principles

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

Abstract

Our thesis in *Fairness versus Welfare* is that social policies should be assessed entirely on the basis of how they affect individuals' well-being. This claim implies that no independent weight should be granted to deontological principles. We support our thesis with three sets of arguments: a demonstration that deontological principles lead to perverse reductions in welfare, indeed, sometimes to a decline in everyone's well-being; the presentation of numerous other difficulties with the principles, including their lack of intellectually satisfying rationales; and a reconciliation of the intuitive appeal of the principles with our thesis that they should not be viewed as directly relevant to the assessment of social policy. In this essay, we explain that the critique of Professor Ripstein largely fails to respond to any of these arguments.

Forthcoming, *Economics and Philosophy* (2004)

---

* Professor of Law, Harvard Law School.
** Samuel R. Rosenthal Professor of Law and Economics, Harvard Law School, and Research Associate, National Bureau of Economic Research. I thank the John M. Olin Center for Law, Economics and Business at Harvard Law School for research support.

Notes on Welfarist Versus Deontological Principles

Louis Kaplow and Steven Shavell

In *Fairness versus Welfare* (*FVW*), we advance the thesis that social policies should be assessed entirely with regard to their effects on individuals' well-being. That is, no independent weight should be accorded to notions of fairness such as corrective or retributive justice or other deontological principles.[1] Our claim is based on the demonstration that pursuit of notions of fairness has perverse effects on welfare, on other problematic aspects of the notions, and on a reconciliation of our thesis with the evident appeal of moral intuitions. Here we summarize our three arguments and explain that Professor Ripstein's commentary largely fails to respond to them. (We will pass over some of what he says because it has little to do with our book, and we will not address his rather surprising attacks on our scholarship because the reader can readily verify their inaccuracy.[2])

---

[1]For a more precise statement of our thesis and how we define "fairness" and "welfare" – including the point that our analysis does not call into question many theories of distributive justice and an explanation of important indirect ways in which fairness may be relevant because it affects welfare – see *FVW* (ch. II).

[2]We find Ripstein's criticisms of our scholarship to be almost inexplicable because most are self-evidently false. Consider, for example, his first extensive attack, in the paragraph and notes on pages 12-13: (1) He states that "E.O. Wilson is quoted out of context (71, n.107; 360, n.137)." (Ripstein, p. 12 n.19.) In fact, E.O. Wilson is not cited in either footnote to which Ripstein refers! Martin Daly and Margo Wilson's book, *Homicide*, is. (2) Ripstein states that we take Korsgaard and O'Neill "out of context" by invoking them "as authority for the proposition that Kant evaluates duties from the standpoint of welfare economics." (Ripstein, p. 12 & n.19). But on the page in question (*FVW* p. 42), we describe Kant as a "strictly deontological philosopher[]." In a footnote (n.55), we quote Kant (twice) to this effect. Our references to Korsgaard and O'Neill appear only in qualifications that follow in that footnote. For example, the first quotation (Korsgaard) is introduced by the statement: "Kant scholars have identified important inconsistencies in Kant's writing that raise some questions about the conventional interpretation" (i.e., the one we adopt). (3) Ripstein insists that we misinterpret corrective justice theorists, treating them as though they ground their views in the need for punishment rather than in the relationship between the parties. Specifically, he asserts that "*[a]fter* the repeated reference to corrective justice theorists," we refer to a punishment-based notion of fairness, whereas, as Ripstein explains, a proper understanding reveals that the relevant theorists "always focus on the *transaction between the parties*," not that the victim suffered or that "the injurer deserves to suffer." (Ripstein, p. 13.) In fact, the passage Ripstein quotes, from page 87, *precedes* our introduction of corrective justice on pages 88-89 and is independent of that subsequent discussion. Thus, we begin our discussion of corrective justice by presenting it as a principle "*[i]n addition* to notions of fairness that might independently call for punishment or compensation"; it is "*another* fairness principle." Furthermore, we immediately describe corrective justice as a principle under which "*the relationship between injurer and victim*, both when harm is inflicted and when compensation is paid, *is critical*," making it difficult to understand how he can claim that we completely fail to appreciate this aspect of the principle.

1. THE CONFLICT BETWEEN FAIRNESS AND WELFARE

Our first theme is that pursuit of notions of fairness results in a pernicious reduction in individuals' well-being. This theme is developed through detailed, sustained examinations of the conflict between fairness and welfare in important contexts – notably, injurers' obligations to victims, contract, procedural rights, and punishment – and with regard to prominent notions of fairness. We identify when conflicts between fairness-based and welfare-based policy recommendations arise in well-articulated paradigmatic situations, examine what if anything might warrant the sacrifices of individuals' well-being, and find justifications for such sacrifices to be lacking. Ripstein's commentary does not seriously address most of this analysis even though it constitutes the analytical heart of our book.[3]

An important aspect of our analysis is the demonstration of the specific further claim that advancing any notion of fairness sometimes makes literally everyone worse off. This strong claim is of evident significance, and not only from a welfarist perspective. For example, those who endorse notions of individual autonomy (or freedom, respect, or integrity) should find it problematic that their principles sometimes require one to endorse outcomes that would be unanimously opposed by individuals, for then everyone's autonomy would be overridden in the name of autonomy.

Ripstein (pp. 6-11) addresses one of our arguments establishing this claim. Namely, we show that when individuals are identically or symmetrically situated, everyone must be made worse off whenever well-being is sacrificed in pursuit of a welfare-independent normative

[3]Ripstein (pp. 16-18) does offer numerous brief (and familiar) criticisms of welfarist (or broadly consequentialist) theories, such as his reference to acceptance of punishment of the innocent (Ripstein, p. 17), but even then he ignores our pertinent discussions, in this case an extensive assessment of the issue (*FVW*, pp. 336-52).

principle.[4] This result holds because, if welfare is ever sacrificed, at least someone's well-being is reduced; furthermore, because everyone is identically affected, it must be that everyone is made worse off.

Ripstein's objections to our use of the symmetric case reflect two basic misunderstandings. First, he is disturbed by the unlikelihood or unreality of cases in which everyone is identically situated. But this point – one that we identify when presenting our argument (*FVW*, pp. 55-56) – is irrelevant. If a principle fails in a basic case in the core of its domain, then the principle is deficient. As Rawls (1980, p. 546) states, "a theory that fails for the fundamental case is of no use at all." Moreover, as we further explain in *FVW* (pp. 57-58, 111-12), the symmetric case is of particular importance for those who endorse the Golden Rule, Kant's categorical imperative, or veil-of-ignorance constructs for evaluating normative principles. The reason is that the use of each construct can, upon analysis, be shown to be tantamount to a requirement that normative principles be assessed *as if* individuals are in a symmetric setting, because some such setting must be envisioned in order to ensure impartiality.[5] Thus, since all notions of fairness do indeed fail in the symmetric setting – specifically, following any of them makes all individuals worse off *whenever* there is a conflict with welfare – they accordingly fail the tests of the Golden Rule and these other constructs.

Second, Ripstein believes that symmetric cases with identical individuals cannot capture important domains of concern because many fairness principles (such as corrective justice) involve asymmetrically situated individuals who have different roles, e.g., a wrongdoer and a

---

[4]Ripstein (pp. 12-16) follows this discussion with another concerning our uses of and misunderstandings concerning private law. Setting aside its inaccuracy, we fail to see its relevance to any of the arguments in our book; none are mentioned. Indeed, Ripstein intermittently notes that we make no claims characterizing private law, a point that we emphasize (*FVW*, pp. 89-90) when we begin our discussion of the corrective justice literature.

[5]Ripstein (p. 8) rejects our use of the word "impartial" in describing a symmetric setting, yet we follow standard practice (and dictionaries) quite literally in using the term to mean "not partial or biased, treating or affecting all equally."

victim. But, as our book explains, one can readily create symmetric cases in such instances simply by positing (keeping in mind that these cases are purely hypothetical) that individuals each play all roles an equal portion of the time. Thus, in the accident context, we suppose that each individual is on one occasion a prospective injurer and on another a prospective victim. (As one can see, anything one might do unto others may, on the other occasion, be done to oneself.) Likewise, in examining contracts, one might assume that each individual is once a buyer and once a seller. If one were to consider whether "might makes right," each individual would in one instance be strong and in another weak. It is obvious that any setting, real or imaginary, may thus be transformed into a symmetric case. Moreover, in our book we actually construct symmetric settings – and ones in which the pertinent fairness principles remain fully operative – in each instance in which we make a symmetric-case argument. Thus, there need be no concern about whether symmetric cases that capture pertinent features are possible to construct.

Ripstein (p. 11) does refer to one of our constructions, but his depiction of what we say is false. Specifically, regarding our discussion of cases involving prospective injurers and victims, he asserts that "accidents are assumed to be inevitable" in our analysis and thus it is inapposite to fairness principles. Yet the first sentence of the pertinent section of *FVW* (p. 99) refers explicitly to "*potential* injurers [who] undertake an activity that *may* cause harm to *potential* victims." Lest there be any doubt, in our actual analysis of each of four different scenarios (pp. 101-104, 118-120, 124-31, 132-33), harm is not inevitably caused. Instead, the occurrence of harm depends on individuals' voluntary decisions whether to act in various ways. And these decisions, in turn, are affected by legal rules – including the negligence rule, which, of course, is fault-based. Hence, Ripstein's characterization is contradicted at every point in our discussion of accidents.

Furthermore, we observe that there is an important gap in Ripstein's overall consideration

of our claim that adherence to any notion of fairness sometimes requires making everyone worse off. Namely, although Ripstein (pp. 6, 9) twice refers to our having formally demonstrated this claim, he does not inform the reader that this formal demonstration (Kaplow and Shavell 2001) does *not* employ a symmetric construct (which is the domain of all of his criticism in this regard). Indeed, when Ripstein (p. 9) quotes us on this matter (*FVW*, pp. 52-53), he excludes the earlier part of our passage that clearly indicates this fact. Thus, even granting all that Ripstein says regarding our symmetric case construct, our conclusion would be unaffected.

Finally, as noted at the outset of this section, Ripstein ignores the bulk of what we actually do in chapters III-VI, where we show that the actual sacrifices in well-being required by adherence to various notions of fairness cannot be justified. Furthermore, much of that analysis is wholly independent of our symmetric-case construct. Notably, our longest chapter, on punishment and retributive justice, does not focus on symmetric settings and barely mentions our argument that everyone may be made worse off by pursuit of notions of fairness, and our assessment of corrective justice and other fairness principles in the accident context independently considers asymmetric cases (*FVW*, pp. 118-23, 132-33), offering further arguments, none of which are mentioned by Ripstein.

2. ADDITIONAL PROBLEMATIC ASPECTS OF NOTIONS OF FAIRNESS

The second theme in *FVW* concerns largely internal deficiencies in notions of fairness. There are serious problems of definition, sometimes leaving basic statements highly incomplete or entirely empty. For example, corrective justice commands rectification by wrongdoers but fails to supply a theory of wrongdoing, without which policy questions cannot be resolved.

Retributive justice requires punishment in proportion to the gravity of the wrongdoer's bad act, but little is said about how to measure the degree of wrongfulness or to determine the correct proportion.

Second, and most important, is the lack of affirmative rationale for notions of fairness. Regarding retributive justice, for instance, scholars from Aristotle to Kant and Hegel to the moderns offer little more than what seem to be conclusory metaphors, such as the need to restore a sort of moral balance after a bad act has occurred. In addition, the problems of definition and of justification interact, for it is hard to refine statements of a theory without regard to its justification, and attempts at justification often lack a clear sense of what is being justified (for example, retributive justice asserts a particular relationship between punishment and wrongful acts, but wrongful acts are undefined and most agree that not all wrongs deserve punishment).

Another difficulty is that many notions of fairness entail a peculiarly selective perspective and thus ignore seemingly pertinent aspects of a problem. Retributive justice, for example, implicitly focuses on those who have been apprehended and found guilty, thus overlooking the many – often the vast majority – who commit similarly bad acts but go scot-free. In addition, the nonconsequentialist character of many fairness principles raises questions of internal coherence. For example, due to inevitable human error, any criminal justice system sometimes produces unjust punishment, including of the innocent; by ignoring and thus potentially sacrificing deterrence, the extent of unjust punishment under purportedly just rules could exceed that arising if unjust punishment were meted out directly.

The foregoing are brief suggestions of arguments that we develop at great length in our book regarding each of the leading notions of fairness that we address in each of the four domains that we consider. We find notable Ripstein's choice to ignore even the existence of this

substantial and obviously controversial portion of our book.

3. RECONCILIATION OF THE APPEAL OF FAIRNESS WITH WELFARISM

Our third theme involves addressing how it can be that notions of fairness appeal to our moral instincts and intuitions and yet should not be accorded any independent weight in policy assessment. We begin by observing that each of the particular principles of fairness that we examine has some correspondence to internalized social norms, like keeping promises and holding wrongdoers accountable. Such norms appeal to us both because of their internalization and related social reinforcement and also because of their practical value in guiding our lives. Given these sources of appeal, it seems natural that we are inclined to give the norms weight when we engage in the analysis of policies to which the norms seem applicable. For example, in considering principles of contract, we are likely to be influenced by our norms regarding promises.

We then explore the implications of this relationship between notions of fairness and social norms for how policy analysis should ideally be conducted. First, the fact of our socialization does not per se provide a basis for treating social norms as normative principles of independent weight. Second, as we explore with regard to each of the social norms in question, the underlying justifications for the norms are ultimately functional (for instance, a practice of keeping promises is conducive to overall well-being). Hence, it would be a mistake for a policy analyst to treat such norms as independent principles to be pursued at the expense of well-being.

Moreover, we identify specific, concrete differences in context between the realms for which the norms and related fairness principles were developed – informal regulation of

everyday social intercourse – and the realm with which policy analysts are concerned – formal regulation using apparatus of the modern state, such as the legal system. Not only that, we suggest that the underlying reason for the divergences between the prescriptions of fairness-based and welfare-based policy analysis that we identify throughout the book can in each instance be traced to these very differences in context.

In light of the foregoing, it should be clear that Ripstein's discussion in the second part of his commentary is not responsive to us.[6] He makes assertions such as "the same kind of argument is equally available, and equally unconvincing, from every direction," (p. 23) and that our mode of argument "is a game at which everyone can play" (p. 35). More broadly, Ripstein writes as if we merely offered a generalized assertion of the possibility that socialization might explain some views, in which case it can as easily be asserted that socialization might also explain other views. But as the foregoing makes apparent, he mischaracterizes what we actually do. Our argument has many steps that Ripstein ignores, and these steps are not reversible at a whim. Moreover, we spent approximately fifty pages (*FVW*, pp. 134-48, 203-13, 241-45, 262-64, 352-72) – above and beyond our initial, extensive statement of our general claim – documenting that each of the aforementioned steps of our argument holds with respect to each of the notions of fairness that we consider. This includes explanations of how differences in context explain the divergences between fairness and welfare with regard to policy prescriptions. For example, in our analysis of the legal process, the primary source of conflict between fairness and welfare in examining civil suits concerns high litigation costs that do not have a significant

[6]Ripstein also, as elsewhere, affirmatively misrepresents our presentation. As but one indication, he (p. 21) chides us for ignoring that our argument is related to arguments not only of Mill and Sidgwick "but also . . . such 20th-century utilitarian moralists as R.M. Hare and J.J.C. Smart." Yet in the very first paragraph introducing our overview of our argument (*FVW*, p. 63), we identify Hare by name in the text and go on to cite him numerous times throughout our discussion. Smart and other twentieth-century philosophers are also cited multiple times.

analogue with regard to informal interaction. Neither the existence of this sort of analysis nor, a fortiori, any of its particulars, is acknowledged or engaged by Ripstein.

4. CONCLUSION

Unfortunately, none of the three major themes that we develop in *Fairness versus Welfare* are accurately represented in Professor Ripstein's commentary. The second is ignored altogether, and the first and third are described in an incomplete and misleading manner. Furthermore, because Ripstein's objections do not relate to what we actually argue and, it turns out, are responded to directly by what we do say, the merit of our thesis is unaffected by Ripstein's remarks. We respectfully urge readers who are interested in the age-old conflict between fairness and welfare to examine our book directly so that they can judge for themselves.

REFERENCES

Kaplow, Louis and Steven Shavell. 2002. *Fairness versus Welfare.* Harvard University Press.

Kaplow, Louis and Steven Shavell. 2001. Any non-welfarist method of policy assessment violates the Pareto principle. *Journal of Political Economy*, 109: 281-86.

Rawls, John. 1980. Kantian constructivism in moral theory. *Journal of Philosophy*, 77: 515-72.

Ripstein, Arthur. 2004. Too much invested to quit. *Economics and Philosophy* (this issue).