Michigan Law Review

Volume 90 | Issue 6

1992

The Quest for Justice

James S. Fishkin
University of Texas at Austin

THE QUEST FOR JUSTICE

*James S. Fishkin**

AFFIRMATIVE ACTION AND JUSTICE: A PHILOSOPHICAL AND CONSTITUTIONAL INQUIRY. By *Michel Rosenfeld.* New Haven: Yale University Press. 1991. Pp. vii, 373. $30.

Two decades into the active revival of liberal political philosophy, a revival that can be dated from the publication of John Rawls' *A Theory of Justice* in 1971,[1] Michel Rosenfeld[2] offers an ambitious account of why all the rival approaches are defective and why a proposal of his own, built around a combination of theories from Lawrence Kohlberg and Jürgen Habermas, provides the ultimate solution to the theory of justice. He then goes on to apply this apparatus to the most vexing issue of current public debate: affirmative action.

Two main questions arise in attempting to evaluate this work. First, how adequate is Rosenfeld's theoretical framework compared to all the rivals he proposes to dispatch? Second, does his theoretical framework support the position on affirmative action he would have us adopt?

Rosenfeld divides liberal theories of justice into libertarian, egalitarian, contractarian, and utilitarian approaches. While this classification does succeed in capturing the main alternatives, it is not intellectually tidy. *Contractarian* refers to a decision procedure, either an actual or a hypothetical social contract, employed as an apparatus of justification. The terms *egalitarian, utilitarian,* and *libertarian* refer to substantive positions rather than to the decision procedure supporting those substantive positions. The difficulty is that prominent examples of substantive positions fall in the egalitarian, utilitarian, or libertarian categories that are supported by a social contract method of justification. Hence Rosenfeld's classifications, which organize a large portion of the discussion, produce more confusion than clarity.

More specifically, Rawls' original position, which bids individuals to evaluate social institutions from behind a hypothetical veil of ignorance, has become the paradigmatic contractarian theory. Yet, as Rawls admits, one can easily deploy a variation of the original position

* Darrell K. Royal Regents Chair in Ethics and American Society, Professor of Government, Law, and Philosophy, University of Texas at Austin. B.A. 1970, Yale College; Ph.D. (Political Science) 1975, Yale University; Ph.D. (Philosophy) 1976, Cambridge University, England. — Ed.

1. JOHN RAWLS, A THEORY OF JUSTICE (1971).

2. Professor of Law, Yeshiva University.

in support of utilitarianism.[3] Would a Harsanyi-style advocate of average utilitarianism (justified by a gambling argument from the original position) fit into the contractarian or the utilitarian category?[4] The strategy of justification is contractarian, but the substantive conclusion is obviously utilitarian.

The matter becomes even more confused by Rosenfeld's adoption of Douglas Rae's categories for understanding the range of "egalitarian" theories.[5] Within Rae's categories, Rawls' "maximin" theory of justice is a species of equality.[6] Is Rawls' theory then a contractarian theory because it is based on the original position, or is it an egalitarian theory because it is advocating a form of equality (at least according to the categories adopted by Rosenfeld)? Furthermore, if social contract theories are characterized by a decision process under actual or hypothetical circumstances, then it is arguable that Robert Nozick's theory, offered by Rosenfeld as the paradigmatic libertarian theory, is also properly classified as a social contract theory. Nozick takes Lockean assumptions to a hypothetical state of nature and shows how a nonrights-violating decision process in that state of nature would yield a minimal state.[7] On some grounds, such a libertarian theory is social contract; in terms of its substantive principles, however, it is libertarian. These categories seem to support classification of Nozick's theory as both social contract and libertarian, Harsanyi's theory as both social contract and utilitarian, and Rawls' theory as both social contract and egalitarian. The truth is that most of the contemporary revival in theories of justice is based, ultimately, on variations of the social contract. To separate egalitarian, utilitarian, and libertarian approaches from the social contract is simply to cut off substantive positions from the strategy of justification most commonly used to support them.

Rosenfeld's own proposal is based on "stage six," a decision procedure that he borrows from the late Lawrence Kohlberg. Kohlberg explicitly classifies Rawls' original position as the best example of stage six.[8] Indeed, Kohlberg says arguments from the original position and those from stage six are substantively equivalent.[9] Yet Rawls' original

3. RAWLS, *supra* note 1, at 12, 163-65.

4. As Harsanyi shows, and Rawls acknowledges, if one takes expected value as the measure of self-interest from behind the veil of ignorance, then one maximizes choice behind the veil by choosing average utility rather than Rawls' proposal. For more on this "gambling argument," see JOHN C. HARSANYI, RATIONAL BEHAVIOR AND BARGAINING EQUILIBRIUM IN GAMES AND SOCIAL SITUATIONS 48-51 (1977).

5. DOUGLAS RAE ET AL., EQUALITIES 12-13, 15-19 (1981).

6. *Id.* at 128-29.

7. ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA pt. 1 (1974).

8. Lawrence Kohlberg, *Justice as Reversibility, in* 5 PHILOSOPHY, POLITICS AND SOCIETY 257 (Peter Laslett & James Fishkin eds., 1979).

9. *Id.* at 263.

position is the paradigm case of a social contract argument. Rosenfeld somehow believes that he can reject the category of social contract arguments and offer us stage six methods as an alternative.

Stage six is the end point in Kohlberg's sequence of empirical stages of moral development.[10] It consists in what Kohlberg calls "ideal role taking" or "moral musical chairs."[11] Rosenfeld quotes this explanation of stage six from Kohlberg:

> a) The decider is to successively put himself imaginatively in the place of each other actor and consider the claims each would make from his point of view.
>
> b) Where claims in one party's shoes conflict with those in another's, imagine each party to trade places. If so, a party should drop his conflicting claim if it is based on nonrecognition of the other's point of view.[12]

Kohlberg terms this process "moral musical chairs":

> Moral musical chairs means going around the circle of perspectives involved in a moral dilemma to test one's claim of right or duty until only the equilibrated or reversible claims survive .... In moral chairs there is only one "winning" chair which all other players recognize if they play the game, the chair of the person with the prior claim of justice.[13]

The key to moral musical chairs is "reversibility." As Rosenfeld explains: "Reversibility involves not only recognizing that others have their own perspective, but also trading positions with others to become aware of the nature and content of their perspectives, so that each gains a richer understanding of the other's objectives" (p. 249). Kohlberg's proposal is a kind of multiperson version of the Golden Rule with the proviso that the claims that are upheld are the ones that would survive each person imaginatively switching places with all the others affected by the choice. We imagine this process continuing until agreement emerges among all the parties.

Kohlberg gives a relatively straightforward example on which Rosenfeld also relies.[14] Kohlberg's moral reasoning studies made use of the so-called "Heinz" dilemma, the story of a woman with a seemingly incurable disease. A "druggist" in the same town has, however, invented a cure for which he demands an exorbitant price. Heinz, the woman's husband, is unable to meet the druggist's demands and steals the drug in order to save his wife's life. Kohlberg cites an interview with an unnamed "Philosopher 3" as exemplifying stage six reasoning:

> Philosopher 3 is saying "start with the Golden Rule, change places with

10. *See, e.g.*, Lawrence Kohlberg, *From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development, in* COGNITIVE DEVELOPMENT AND EPISTEMOLOGY 151 (Theodore Mischel ed., 1971).

11. Kohlberg, *supra* note 8, at 262.

12. P. 249 (quoting Kohlberg, *supra* note 8, at 263).

13. P. 249 (quoting Kohlberg, *supra* note 8, at 262).

14. P. 250 (citing Kohlberg, *supra* note 8, at 259).

> the wife in deciding. Is your denying a duty to save the woman's life consistent with the Golden Rule?" The wife, we may say, holds that her right to life is higher or prior to the druggist's right to property. She claims that the husband has a duty to steal to protect this right, since she cannot. The druggist denies that the husband has a duty to steal the drug and asserts he has a right to property equal to or greater than the wife's right to life.[15]

The test is whether, when the druggist imaginatively changes places with the wife, he can uphold the claim that the property rights are more important than the human life:

> Is the druggist's denial of the husband's duty to steal reversible? No, this denial could not stand if he exchanged places with the wife. In the position of the druggist, he holds his right to property higher than the wife's right to life. Presumably, however, if it were his life at stake, not the wife's the druggist would be rational enough to prefer his right to life over his property and would sacrifice his property. If the druggist tried to make his conception of rights and duties reversible by imaginatively changing places with the wife, he would give up the idea that the husband had a duty to respect his property rights and would see that the husband had a duty towards his wife's life.[16]

Kohlberg offers this as a simple paradigm case of stage six reasoning. Rosenfeld encounters more difficulty, however, when he applies this method to his difficult case of affirmative action. Rosenfeld explicitly wants to "concentrate on the most controversial practices" (p. 47). For Rosenfeld "*affirmative action* shall be assumed henceforward to include some kind of preferential treatment" (p. 47). The controversy comes in because "preferential treatment shall include the hiring or promotion of a minority or woman over a more qualified nonminority or male" (p. 48).

Apply this kind of case to the same stage six reversibility test Kohlberg uses for the Heinz dilemma. If a "more qualified nonminority or male" switches places in moral musical chairs with a less qualified woman or minority male, it is hard to imagine that the nonminority or male will come to agree that the interests of the less qualified minority should override his interests. In the more limited circumstance where the minority or woman applicant were an actual victim of discrimination or came from a severely disadvantaged background, then we might be able to imagine a dialogue in which the nonminority or male would agree, when the two switched places, to a policy of preferential treatment. But Rosenfeld's principle is not defined in this narrow way. It yields a kind of affirmative action that explicitly includes the most controversial cases, namely those where a less qualified minority, regardless of class background or income tests, is preferred over a more qualified nonminority or male. Whatever the

15. Kohlberg, *supra* note 8, at 261-62.

16. *Id.* at 262.

merits of such a policy, it is hard to imagine it emerging consensually from the role-switching of moral musical chairs.

At some points Rosenfeld appears to invoke group rather than individual comparisons to support such a policy. For example, he constructs a dialogue with the "Innocent White Male" who may lose out from preferential treatment and concludes, "the more pervasive racism and sexism are in a given social setting, the more it would seem that affirmative action could play a key role in the transition of blacks and women from a status of inferiority to one of equality" (p. 325). If one is concerned with the general status of one group compared to another, then the preferential treatment argument takes on a different force. Yet it is difficult to see how one is to square such a concern for *groups* (defined by racial or gender categories) with the quite specific apparatus of reversible reciprocity or moral musical chairs involving *individual* role-switching. Rosenfeld admits as much when he adopts, at the beginning of his book, an individualistic restriction on admissible arguments. In endorsing a general postulate of equality as a presupposition for the argument that follows, he tells us that "it is the individual rather than the group who is entitled to such equality" (p. 4). The result is that "purely group-related justifications of affirmative action become unacceptable" (p. 4). More specifically, "a group proportionment argument advocating affirmative action to eliminate glaring group-related imbalances in a particular workforce would have to be rejected as inconsistent with the conviction that the individual rather than the group is the proper subject of equality" (pp. 4-5). The difficulty is that, within this individualistic postulate, the basis for preferential treatment of a group, as such, becomes quite tenuous. If Rosenfeld were to abandon his individualistic version of liberal theory, he could get the controversial conclusion he seeks. But he seems to want liberal individualism in theory and group preferences in practice, somehow believing that the latter can be made to follow from the former.

Rosenfeld introduces an additional theoretical apparatus that might be intended to help at this point. He says Habermas' notion of the "ideal speech situation" provides "a critical tool designed to expose communication distortions and to indicate the directions in which a normatively oriented reconstructed dialogue relying on undistorted communication would lead" (p. 259). The idea is that some preferences result from self-deception or false consciousness and can be ruled out of the dialogue to which Kohlberg's reversible reciprocity or moral musical chairs applies. "[W]hat prevents one person from being able to appreciate the perspective of another may frequently be traceable to differences in power and ideology or to instances of deception and self-deception" (p. 260). We are supposed to reconstruct the relevant preferences by reference to what we would imagine being upheld in a Habermasian ideal speech situation so as to rule out these

distortions. Rosenfeld gives the example of a black person objecting to the justice of affirmative action:

> That black participant would be the victim of self-deception, however, if the actual reason for his claim were that he believed that because of his superior talents the availability of affirmative action would tend to devalue his performance in the eyes of whites, thus reducing his ultimate chances of success in a white-dominated society. [pp. 262-63]

We are not given any further explanation as to why such a conclusion would have to be the result of "self-deception." It would seem, in fact, to be an empirical issue whether, in a particular context, affirmative action had this effect. Rosenfeld does not offer empirical evidence on the issue. Nor does he offer the kind of counterfactual dialogue that would give plausibility to the claim that the black man's conclusion could not be upheld in a Habermasian ideal speech situation.

Just as the basis for inferring self-deception is a bit mysterious in this example, it is difficult to see how the conclusions of our white male in the affirmative action dialogue should be revised. Once more, there might be a basis for such conclusions if we were comparing someone from an actually disadvantaged background with a privileged white male. But when affirmative action is applied in competitive meritocratic contexts, based on racial or gender categories per se, there is the well-known problem that comparatively well-off members of the group will commonly be beneficiaries.[17] If such members are to be hired even when they are less qualified (as would be required by Rosenberg's definition of preferential treatment cited above) then the most controversial version of affirmative action would have been implemented. Once again, it is hard to see how such conclusions can result from "reversible reciprocity" even when Habermas is invoked to rewrite some of the preferences that can be attributed to some of the participants.

Rosenfeld offers one other strategy for resolving such disagreements. He stipulates that when priorities among the conflicting claims do not emerge from the reversal of perspectives test, and when the intensity of feeling on each side of the issue is comparable, then the issue can be resolved by utilitarian calculations as to the balances of net preferences on the issues to be decided. He offers the example of a municipality's deciding between a swimming pool and tennis courts.

> Under these circumstances, if each resident switched places with every other resident, they would all realize that individual preferences for a swimming pool are no more intense than those for tennis courts. Accordingly, each resident would be justified in concluding that the preferences of others are not entitled to any greater deference than her own preference.

17. JAMES S. FISHKIN, JUSTICE, EQUAL OPPORTUNITY, AND THE FAMILY 103-05, 121-30 (1983).

> . . . justice as reversibility would arguably be satisfied by a resolution of conflicts among such claims on the basis of calculations of net preferences or net utilities. [pp. 257-58]

It is not clear why the balance of feeling about the issue should have become a criterion for evaluating the merits of the claims or interests at stake on each side. Notorious difficulties plague the interpersonal comparison of intensities and the use of felt intensity as a criterion of interests.[18] Rosenfeld does not air these issues but simply stipulates this strategy. Furthermore, it is hard to see how Rosenfeld's approach could be employed to settle the issue of preferential treatment. Both the candidate who is nonminority or male and the candidate who is minority or female or both are likely to feel equally intensely. However, if utilitarian calculations involving everyone's preferences were then invoked to settle the issue (in the manner Rosenfeld imagines for the swimming pool example), it is far from clear that preferential treatment would emerge. Preferential treatment does not enjoy widespread public support in opinion polls in this country.[19] In any case, this last strategy would render the issue dependent on the distribution of preferences that happened to be prevalent at a given time throughout the country. Whichever strategy Rosenfeld employs, it hardly seems that he has provided a more determinate basis for settling this controversial issue than have rival philosophical approaches.

In particular, it seems doubtful that Rosenfeld's effort to combine Kohlberg with Habermas can be taken, as he believes it should be, to constitute a theoretical strategy superior to Rawls' theory of justice. Rosenfeld objects to the Rawlsian approach because "the effect of the veil of ignorance is to remove from each individual that which makes his or her own perspective different from that of others" (p. 250). Rosenfeld concludes that behind the veil of ignorance, "the elimination of all differences makes reversibility essentially trivial" (p. 250).

A Rawlsian might reply that the differences are eliminated in the original position because they are held to be irrelevant from a moral point of view. Behind the veil of ignorance I am not supposed to know my race or my class or my I.Q. because if I did, I could tailor principles to the advantage of people with those characteristics. Because the Kohlbergian version permits us to know about them, these differences may upset the emergence of any consensus from the moral decision procedure. Kohlberg has a relatively easy time when the issue is the

---

18. For my previous overview of the problem, see JAMES S. FISHKIN, TYRANNY AND LEGITIMACY 20-25 (1979).

19. In the 1990 national election study, respondents were asked a carefully balanced question that ended with "are you for or against preferential hiring and promotion of blacks?" Respondents in favor numbered 237 of 1007, or 23.54%, while respondents against numbered 698 of 1007, or 69.31%. Warren E. Miller et al., American National Election Study 1990: Post-Election Survey (1991) (on computer file at Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan).

life of Heinz's wife versus the profits of the druggist. But the system has more to overcome when the issue is the claim of a more qualified nonminority versus that of a less qualified minority — the problem Rosenfeld faces in his chosen controversial topic of preferential treatment.

Despite these difficulties, there is the basis for an important insight in this book: it attempts to combine dialogue among actual people with the impartial perspectives of hypothetically rational choice. One important strand of liberal theory, actual consent, would emphasize real choices of real people. The actual consent of the governed binds in the way that a promise does. But actual consent, apart from its notorious ambiguities,[20] also has the defect that we may, sometimes rightly, suspect the socialization and rationality that produces the consent. To take an obvious case, if citizens are brainwashed or indoctrinated, we hardly view their conclusions as morally binding or legitimate.

By contrast, the revival of liberal theory has been based on hypothetical thought experiments that attempt to reform the biases of socialization through some impartial decision procedure. These respond to the deficiencies in socialization and indoctrination through some mechanism requiring an impartial perspective. Rawls' veil of ignorance is only the most celebrated example. Ackerman's neutral dialogue,[21] the perfectly sympathetic spectator of the classical utilitarians,[22] and the gambling argument for average utility provide some other examples. The difficulty is that these mechanisms offer competing hypotheticals. Each offers a rival hypothetical choice procedure, but there is no actual commitment to any one of the competing candidates and no metaprocedure for deciding among the claims of the competing hypotheticals. They all make a putative claim to fairness or impartiality. Furthermore, as the debate between Rawls and Harsanyi demonstrates, very slight alterations in the statement of the assumptions can lead to radically different principles. In that case the difference between assuming, behind the veil of ignorance, an equal chance of being any given person, and mere uncertainty about the likelihood of being any given person is enough to support average utility rather than maximin justice (requiring maximization of the minimum share of primary goods).[23]

Actual and hypothetical choice strategies each have advantages and disadvantages in the theory of justice. If a theory were somehow

20. For a good discussion, see Hanna Pitkin, *Obligation and Consent, in* 4 PHILOSOPHY, POLITICS AND SOCIETY 45, 51-62 (Peter Laslett et al. eds., 1972).

21. *See* BRUCE ACKERMAN, SOCIAL JUSTICE IN THE LIBERAL STATE (1980).

22. For a good brief discussion, see RAWLS, *supra* note 1, at 183-88.

23. For an extended discussion of these hypothetical choice strategies and their controversial claim to impartiality, see JAMES S. FISHKIN, THE DIALOGUE OF JUSTICE (forthcoming 1993).

to have the bindingness of actual commitments and the impartial character of recent hypothetical procedures, then a marriage of the two approaches might be more powerful than either taken singly. In significant ways, this is the attraction of what Rosenfeld tries to do with his combination of actual dialogue and impartial procedure. These criticisms, while focused on several particulars, should not obscure the promise of the general strategy. Actual agreements — actual choices — should be taken as binding provided they can be insulated from delegitimating charges that they result from coercion, brainwashing, or indoctrination. The theory of justice needs to be brought down to earth from imaginary thought experiments. Rosenfeld only charts a somewhat bewildering path, but his book points in the right direction.