RAWLS ON JUSTICE AS FAIRNESS*

James M. Buchanan
Virginia Polytechnic Institute and State University

When I first encountered John Rawls' conception of "justice as fairness," I was wholly sympathetic. I interpreted his approach to be closely analogous, even if not identical, to that aimed at explaining the voluntary emergence of "fair games," with widely divergent applications. Stimulated by Frank Knight and, more directly, by Rutledge Vining (both economists who worked independently of and prior to Rawls), I sensed the possible extensions in the explanatory-descriptive power of models for "rules of games," derived in accordance with some criteria of "fairness." As readers of *The Calculus of Consent* recognize, Gordon Tullock and I employed such models in our derivation of the "logical foundations of constitutional democracy" (our subtitle), of a political structure not grossly divergent from that envisoned by the Founding Fathers and embodied in the United States Constitution, at least in its initial conception. For these, and other reasons, I looked forward to publication of Rawls' long-promised treatise.

Now that the book has appeared, I find myself less sympathetic with Rawls than I might have anticipated from my early reading of his basic papers. There are two distinct reasons for this temporal difference in assessment, and this review incorporates a two-part argument. Rawls has extended his allegedly contractarian conception and thereby increased its vulnerability. On closer examination, Rawls does not seem to say what I thought he was saying. His approach now appears quite different from that which I shared in 1960.

The second reason for a change in my own reaction to Rawls' work lies in a shift in my own thinking since 1960. I am less of a contractarian, although just where my own position would now be classified remains an open question.

I.

Consider several persons voluntarily discussing the rules for an ordinary card game which they are to play. No one can predict the particular run of cards on any series of plays or rounds. These persons attempt to agree on a classification scheme that allows them to separate "fair" and "good" games (defined by sets of rules) from "unfair" and "bad" games. Agreement requires predictions about the working properties of alternative sets of rules, but since these predictions are at least

* A review article on John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971).

quasi-scientific in nature, there is no self-interest barrier to consensus. For example, a set of rules that insures victory in all subsequent rounds of play to the chance victor of the intitial round might be labeled "unfair" and "bad" and the game embodying such rules rejected out of hand. By contrast, a set of rules that guarantees independence among opportunities over separate rounds of play might be classified as "good" and "fair." Or, alternatively, rules that penalize the victor in one round of play over a finite series of subsequent rounds might equally qualify by criteria for "fairness" and "goodness." These simple examples suggest that there may be many sets of rules, many games, that would meet reasonable criteria of both "fairness" and "goodness." As among these, there seems no apparent means of selecting a single "best" set.

It is surely reasonable to extend this essentially contractarian framework to an evaluation and analysis of the whole set of rules, formal and informal, that describe or might describe social interaction. The framework appeals quite naturally to anyone who accepts the individualistic or Kantian precept that human beings are to be treated as ends never as means. This precept implies, in some basic sense, that men are to be treated "as equals." The appropriate question becomes: How would a group of individuals, no one of whom can predict his own position in any of the time segments over which the rules to be chosen are to be operative, go about setting up the socio-political rules of the game? The "veil of ignorance" or uncertainty is the device or requirement that forces participants to consider alternatives on grounds other than identifiable self-interest, narrowly interpreted. In a broader sense, of course, the objective of individual self-interest is served precisely by the criteria for "fairness" and "goodness" in the rules, in the fundamental constitution of society. The important usage of this framework is to evaluate and to analyze existing and proposed social institutions.

It is possible to exclude some existing institutions and some proposed institutional-constitutional changes on such minimal contractarian grounds. Overtly discriminatory restrictions on the franchise, for example, clearly violate the precept of equality among participants. Similarly, the criteria might rule out restrictions on entry into professions. This essentially negative application of the "fairness" criterion can be helpful. But can positive application be other than classificatory? As the simple game examples above suggest, there may exist a whole set of socio-political institutions, embodying among themselves quite different internal characteristics, that qualify on the minimal criteria of "fairness" and "goodness."

Rawls' first principle for a "just" social order is that of equal liberty for all persons. It seems plausible to suggest that any departure from this principle would be rejected in any system that qualifies on the minimal criteria. But Rawls' second major principle seems on much weaker ground. To his first principle of equal liberty, Rawls appends lexiographically his second "difference principle." This

states, specifically, that allowable distributional inequalities among persons are acceptable only to the extent that their existence benefits the least-advantaged members of the community. I should accept the hypothesis that a socio-political-economic structure embodying the difference principle meets widely-accepted criteria of "fairness." But I should not be prepared to elevate this principle into the ideal position accorded it by Rawls. There may be many other distributional rules that qualify within the acceptable set, classified only by the minimal criteria for "fairness" and "goodness."

Let me illustrate this with an extremely simple numerical example. Two potential players consider alternative positive-sum games, A and B, each of which involves only one round of play. The playoff structure in A is 60:40, while that of B is 80:20. So long as each player has an equal opportunity to win or lose, is there any reason why A should be accepted as "fair" and B as "unfair"? Both games would seem to qualify as "fair" under a broader conception than Rawls would accept.

This is more than a disagreement on detail. By his attempt to make the contractarian approach or model do more than is appropriate, Rawls seems to fall into precisely the same trap as the utilitarians, whom he quite properly criticizes. As he finally admits, Rawls is an idealist, and he seeks to lay down the principles of a "just" social order. He is extremely cautious, and he does allow for much more latitude than most of his idealist colleagues through the ages. But to me he is a bizarre contractarian, despite his self-identification. Perhaps my professional economist's biases show here, but the very essence of contract is the *nonspecification* of outcome by external observers. Traders trade; agreement is reached, agreement that is presumed to be mutually beneficial to the parties. Conceptually at least, there is a subinfinity of possible equilibria along some generalized contract locus. The task of the contractarian social philosopher is to evaluate and to analyze the institutions of the trading process, to lay down criteria for "fairness" in these rules (e.g., that contracts are enforced once made, that fraud is prohibited, that all markets are open, etc.). The task does not involve specifying distributional attributes of outcomes. This unwillingness to allow for a multiplicity of "games," accompanied by a zeal for normative uniqueness, has plagued and continues to plague modern welfare economics. It is singularly unfortunate that Rawls has come so near to what I should classify as a genuine contractarian position while yet remaining so far removed in this most fundamental respect.

Rawls might respond to this criticism by charging that my conception is purely procedural whereas he states explicitly his desire to go beyond procedural limits. In my view, however, there is a direct relationship between a contractarian philosophy of social order and a willingness to be bound by such limits. As contractarian, I cannot, without stepping outside my own limits, lay down precise

descriptions of the "just society" or the "good society." I must abide by my own standards and accept as equally "just" whatever outcome emerges from the negotiations among freely-contracting persons in an idealized "original position," constrained by the "veil of ignorance."

II.

My second reaction is as applicable to my own earlier conception of the contractarian position, implied in the discussion in Part I, as it is to the more vulnerable position espoused by John Rawls. What point is there in talking as if persons will "think themselves" into some idealized version of an "original position" behind a deliberately contrived "veil of ignorance," when we know that, descriptively, the men who must make social choices are not likely to make such an effort? Social choices will continue to be made, as they have always been made, by ordinary mortals, with ordinary passions. Recognizing this, what can be said about ordering rules in terms of criteria for "justice" and "fairness"—whether these criteria be broadly or narrowly drawn? David Hume's stricture that reason must be, and should be, slave to the passions can be helpful here. Precisely because we recognize ourselves to be ordinary men, no different at base from others of our species, we can cultivate an attitude of mutual tolerance and respect for one another, along with a highly skeptical attitude toward anyone who presumes to lay down ideal standards whether or not this is accompanied by proposals collectively to force such standards upon us. But more than this is required. Rules for social order must be evaluated and analyzed, and criteria for orderly change in these rules must be developed. While I can fully appreciate the desire to search for more, the limits seem apparent to me. We can, first of all, emphasize the categorical distinction between "constitutional" choice, the choice from among alternative sets of rules or institutions, and "operational" choice, the choice of policy outcomes within a given constitutional-institutional order. In Rawls' framework, this distinction is not important, and it is not surprising that he largely neglects it. In the more realistic setting that I am suggesting here, the distinction is of critical, indeed crucial, importance. And for reasons closely analogous to those through which Rawls justifies his notion of the original position. In a short-run, operational context, when choices are made within an existing constitutional structure (e.g., a legislative decision on welfare or tax reform), it is folly to expect representatives of the separate consituency interests to act on some idealized principles of "justice." And, indeed, I am not at all sure that we should desire a system where representatives tried to follow such principles, if such were possible. Self-interest can be turned to good account, and even in political process it offers some ultimate protection against ideologues of all stripes.

In the distinct, and conceptually separate, constitutional context, when choices are made among alternative sets of rules, there are at least *some* elements

characteristic of the "veil of ignorance." To the extent that rules are considered as permanent or quasi-permanent, designed for operation over a time sequence that remains perhaps open-ended, individual participants in the selection process must be uncertain as to just where their own self-interest lies. They will, to this extent, be motivated to opt for rules and rules changes that embody "fairness" or which at least contain "unfairness" within broad limits. Certain institutional devices may assist in generating the desired uncertainty here, for example, explicitly-chosen delays in the implementation of choices, legal precedent, prohibitions of personal aggrandizement from political offices, etc. But despite all this, interests are identifiable even over long terms, and men will act to further them. Rules for social order, as observed, will reflect the struggle among interests, and will rarely, if ever, qualify as "just" in accordance with any idealized criteria.

This raises issues of compliance with or adherence to allegedly "unjust" rules. If consensus is attained, such rules may, of course, be changed. But lacking consensus, who is to decide when criteria of "justice" override the stability of law? We start always from here, not from an "original position," and if men have not previously been guided by agreed-on precepts of justice, what expectation can we have that those to whom we might offer power will behave differently? Conservative or reactionary it may be, but attainable consensus offers the only meaningful principle for genuine constitutional change.

How far is this from Rawls' conception? I do not know, and I cannot tell from a careful reading of this treatise. Would John Rawls allow Earl Warren or his successors to make judgments on their own versions of "justice as fairness." and, in the process, to disregard the embodied predictability of existing constitutional order? I wish that I could be sure that Rawls would answer negatively to this question. If he is advancing "justice as fairness" as a basis for discussion, as an input in some process of reasoning together, with consensus as an ultimate objective as well as constraint, I should grant his work high marks indeed. If, however, he is holding up "justice as fairness" as the embodiment of "truth," which judges and legislators in their "superior wisdon" are to force upon us, Rawls' book deserves to gather dust on the idealist bookshelf.

Is not misguided idealism, operating in disregard of constitutional precepts, a major source of our time's tragedy? When the judiciary is allowed to make "constitutional" choices that cannot secure minimally required legislative assent, and when the judiciary is respected and applauded in the process, the misunderstanding of constitutional democracy has indeed gone far. And when the people, acting through the legislative arm of government, even find it so much as necessary to begin to discuss reversing court-ordered constitutional change by the amendment process, the confusion has come full circle.

It is perhaps inappropriate and unfair to expect *A Theory of Justice*, a major philosophical treatise, to enlighten us on matters relevant to modern politics. I wonder. The social order that James Madison tried to secure in the United States Constitution, and which was respected for almost two centuries, did not fully embody "justice as fairness." Through time, however, this constitution was adjusted by both a responsive legislature and a responsible judiciary to move closer toward the satisfaction of Rawlsian or alternative precepts of justice. The emergence of unpredictable legal chaos came as the judiciary began to assume the role as guarantors of "justice" in some idealistic sense. Surely we now need a wider recognition of man's inability at playing God. It is a matter for regret that the extension and elaboration of his basically humble, and, to this degree, admirable conception of "justice as fairness" led John Rawls away from rather than toward the contribution to social philosophy that this treatise might have represented.