

Which Is the Fairest One of All? A Positive Analysis of Justice Theories

JAMES KONOW¹

No man during, either the whole of his life, or that of any considerable part of it, ever trod steadily and uniformly in the path ... of justice, ... whose conduct was not principally directed by a regard to the sentiments of the supposed impartial spectator, of the great inmate of the breast, the great judge and arbiter of conduct.
– Adam Smith (1759) p. 357

1. Introduction

Justice arguments are now widely invoked to improve theoretical and empirical analysis in nearly every field of economics. Incorporated into game theory (e.g., Matthew Rabin 1993), fairness predicts the deviations from pure self-interest observed in many laboratory experiments (e.g., Werner Güth and Reinhard Tietz 1990). Its impact has also been cited in many real-world contexts, including the intermittent failure of product markets to clear (Daniel Kahneman,

Jack Knetsch, and Richard Thaler 1986); resolution of social choice problems such as locating nuclear-waste facilities (Felix Oberholzer-Gee, Iris Bohnet, and Bruno Frey 1997); public-utility regulation (Edward Zajac 1985); and labor unemployment due to efficiency wages (e.g., George Akerloff and Janet Yellen 1990). The view that “By now we have substantial evidence suggesting that fairness motives affect the behavior of many people” (Ernst Fehr and Klaus Schmidt 1999) is expressed in mainstream economics. This contrasts with the traditional belief of many economists that justice is chimerical or amorphous. A more sympathetic stance placed it outside the domain of economics, better left to philosophers, political scientists, or sociologists. There has been a steady trend, however, of increasing interest in and acceptance of justice in the economics profession, even partially displacing efficiency.² This is not to say, of course, that economists are or should be abandoning their traditional

¹ Loyola Marymount University. I thank the editor and three anonymous referees of the *Journal of Economic Literature*; Alison Alter, Gary Bolton, John Coleman, Gary Charness, James Devine, Jon Elster, Duncan Foley, Simon Gächter, Wulf Gaertner, Guillermina Jasso, Serge-Christophe Kolm, Alexander Kritikos, Axel Ockenfels, Joe Oppenheimer, Richard Posner, Matthew Rabin, Erik Schokkaert, John T. Scott, Alois Stutzer, Peyton Young, Ed Zajac, and participants at the meetings of the Public Choice Society, Social Choice and Welfare Society, and International Society for Justice Research for many helpful suggestions and comments. Any remaining errors or shortcomings are, of course, my own. I also thank Jack Knetsch for permission to use questions from Kahneman, Knetsch, and Thaler (1986).

² This is suggested, for example, by an examination of studies documented on *EconLit*. The number of entries for the 1970s under the keyword “efficiency” outnumber those under “justice” or “fairness” (not counting those under the equivocal term “equity”) by sixteen to one. For the 1980s this ratio falls to about nine to one, and for the 1990s this gap further narrows to 4.4 to one. In fact, if one considers entries under the *JEL* classification system in operation since 1991 through the present, hits under the code closest to justice (D63: Equity, Justice, Inequality, and Other Normative Criteria and Measurement) outnumber those under that closest to efficiency (D61: Allocative Efficiency; Cost-Benefit Analysis) almost two to one.

interest in efficiency. Instead, stimulated by empirical evidence and, perhaps, the perception of increasing economic inequality, they are expanding their studies to encompass a wider set of distributive concerns. Despite the emerging consensus in economics over the relevance of fairness, though, no such agreement yet exists among economists or, for that matter, among psychologists, political scientists, sociologists, or philosophers, about the proper theory of justice.

1.1 *Two Goals of the Study*

One goal of this paper is to conduct a *positive analysis* of leading positive and normative theories of justice, where a remarkable lacuna exists in the literature.³ By positive analysis I mean that each theory, whether originally conceived for this purpose or not, will be evaluated in terms of how accurately it describes the fairness preferences of people. In this paper, the terms fairness, justice, and equity always refer to the view of Adam Smith's impartial spectator whose judgment is not biased by any personal stake. The discussion includes both *distributive justice*, which concerns fair outcomes, as well as *procedural justice*, which addresses fair processes, whereby the more extensive treatment of the former reflects the relative emphasis in the justice literature. Justice is operationalized here mostly in relation to material wealth, the chief concern of most economists, even though it is clear that the forces discussed often impact noneconomic domains. Other factors that affect allocations include altruism, reciprocity, spite, kinship, and friendship. These are significant but distinct phenomena, which nevertheless underscore the import and timeliness of studying justice, given growing evidence that some behavior previously attributed to these forces (especially reciprocity) is likely due to distributive preferences.

³ There are, however, excellent surveys on more narrow topics from which this paper has also profited, e.g., Bernard Cullen (1994) reviews *normative philosophical theories* and Erik Schokkaert (1994) *normative economic theories*.

A second, closely related goal of the paper is to propose and defend an *integrated* justice theory that synthesizes previous approaches and explains actual values as the conflation of four distinct forces or *elements*. These elements of justice inspire four corresponding theoretical categories (or families) into which each of the theories is placed and analyzed. The category *equality and need* covers theories that incorporate a concern for the well-being of the least well-off members of society including egalitarianism, social contract theories (chiefly Rawls), and Marxism. They inspire the Need Principle, which calls for the equal satisfaction of basic needs. The *utilitarianism and welfare economics* family comprises utilitarianism, Pareto Principles, and the absence of envy concept, which have grown out of consequentialist ethics, or the tradition in philosophy and economics that emphasizes consequences and end-states. They are most closely associated with the Efficiency Principle, which advocates maximizing surplus. The category *equity and desert* includes equity theory, desert theory, and Robert Nozick's theory. Together they inform the Equity Principle, which is based on proportionality and individual responsibility. The *context* family discusses the ideas of Kahneman, Knetsch, and Thaler; Michael Walzer; Jon Elster; H. Peyton Young; and Bruno Frey and Alois Stutzer, among others. This fourth family does not generate a distributive principle but rather deals with the dependence of justice evaluation on the context, such as the choice of persons and variables, framing effects, and issues of process.⁴

⁴ When dealing with such an extensive literature, even a wide-ranging review cannot be comprehensive. Although I have striven to include the most influential theories of justice, some theories are omitted because they are not primarily theories of justice (e.g., game theories), or because their focus is more remote from the subject matter of economics (e.g., juridical theories), or because their incorporation into the four elements that frame the study seems forced (rights theories). Actually, the paper seeks to represent the breadth of the literature in a relatively concise manner by treating many theories while focusing on those aspects of each that contribute to the integrated theory.

While proceeding through the sometimes intricate analysis that follows, the reader can better maintain a sense of unity if he or she keeps in mind the dual goals of this paper and the framework that structures them. On the one hand, the specific theories discussed offer very different, and sometimes contradictory, perspectives on the meaning of justice. On the other hand, I argue for a general theory of justice as a unifying framework for the specific theories. These ostensibly dissonant objectives are reconciled by the following two facts. First, the general theory guides the classification of a specific theory into the category (i.e., element of the general theory) that is judged as most helpful for distilling the specific theory's most salient contribution to understanding actual justice views. Nevertheless, the evidence, taken as a whole, does not confirm any single theory in toto and sometimes even refutes central suppositions or conclusions. Both favorable and unfavorable evidence on the specific theories, however, produces lessons for the general theory. Second, it should be emphasized that the general framework around which the analysis is organized is an *integrated theory*, but not a *composite theory*: justice is more than the sum of its parts. The three principles of justice must be weighted, and context provides the weighting scheme in specific cases. The argument is that each category captures an element that is important to crafting a positive theory of justice but that no single family or theory within a family suffices to this end. Instead, fairness views are best explained by an integrated approach that acknowledges the influence of the three principles of justice, whereby the weight on each is determined by the context. This method enables one to treat justice rigorously and to reconcile results that often appear contradictory or at odds with alternative theories.

1.2 *Reasons for this Research Agenda*

People justify their positions and behavior in a wide range of situations based on justice, for example, in connection with affirmative

action, global warming, labor-management conflicts, "fair" trade negotiations, and debates on the taxation of income, inheritances, and corporate dividends. The frequency and vehemence of such claims, often accompanied by sacrifices, attest to a conviction on the part of the advocates regarding both their normative value and their power to persuade and, thereby, to alter outcomes. These observations are significant because they indicate that fairness, in fact, appeals to a common moral sense, which, when applied to specific cases, is subject to some interpretation. In particular, biases often emerge when stakes are involved; e.g., Kenneth Binmore (1994) reports a strong tendency by subjects, when debriefed following bargaining experiments, to describe their self-serving decisions during the experiments as "fair." Various studies, including those of Linda Babcock et al. (1995), Tore Ellingsen and Magnus Johannesson (2003), and myself (Konow 2000), trace this bias in large part to deception, both of others and of oneself, regarding what is fair. These studies also indicate that, although biases sometimes widen the range of predicted outcomes, behavior is still constrained by fairness. Thus, justice is not amorphous or arbitrarily malleable, and, as I seek to show in this paper, fairness preferences usually converge when stakes are removed.

These facts suggest at least two important reasons for seeking a descriptively accurate theory of impartial justice. First, social scientists must consider how justice, alone or in tandem with other goals (such as self-interest or reciprocity), affects the phenomena they study. Although stakeholders often subject justice to biased and differing interpretations, in order to have moral force, their claims cannot be capricious but must be constructed around impartial standards. Whereas observed behavior typically results from multiple motives, a study of impartial justice consciously aims at separating the effects of unbiased justice, biased justice, and other motives.

