

Introduction

Unlike lesser thinkers, John Rawls readily accepted criticism and openly made adjustments in his positions, while steadfastly defending what he believed could withstand criticism. Thus, both continuity and change are evident in the thirty-year interval between *A Theory of Justice* (1971) and *Justice as Fairness: A Restatement* (2001).¹ Taken together, his writings comprise the most thorough and sophisticated body of work by an individual in the history of political philosophy. This achievement was enabled by Rawls's careful study of the great work in the tradition that preceded him, by his familiarity with the literature in related scholarly fields such as economics and developmental psychology, and by his involvement in developments in contemporary philosophy more generally – notably, his elaboration of the method of “reflective equilibrium,” which furthered the ascendancy of the now-orthodox Quinean/Duhemian view of theory construction.

Unlike many of the political philosophers who preceded him, and whose work he assimilated, Rawls was not personally caught up in the great social and political upheavals of his day. He served with distinction in the Pacific Theater in the Second World War, but he – unlike a number of his contemporaries – was not a conspicuous spokesperson or actor in the civil rights movement or in the opposition to the War in Vietnam. He is said to have crossed a picket line, but never (so far as I can tell) to have joined one.² Unlike Hobbes and Locke, he did not rely on the

¹ Citations to Rawls's works are abbreviated: TJ = *A Theory of Justice*, 1999 rev. ed.; PL = *Political Liberalism*, 1996 paperback ed.; CP = *Collected Papers*; LP = *The Law of Peoples*; JF = *Justice as Fairness: A Restatement*; GT = *Justice as Fairness: a Guided Tour*; LHPP = *Lectures on the History of Political Philosophy*; LHMP = *Lectures on the History of Moral Philosophy*; BI = *A Brief Inquiry into the Meaning of Sin and Faith*. Full citations appear in the bibliography. In the text, I often refer to *A Theory of Justice* as “Theory,” to *Justice as Fairness: A Restatement* as “the Restatement,” and to *The Law of Peoples* as “Law of Peoples.”

² During the Vietnam War, Rawls was deeply opposed to the Selective Service System's policy of issuing four-year undergraduate student deferments from conscription, “quite apart from the injustice of the war itself” (Pogge 2007, 19–21).

patronage of any powerful person, and was not an advisor to any. Unlike Marx, he was not identified with a party or a movement. Unlike Hobbes, Locke, Rousseau, and Marx, he was never obliged to flee persecution. In many respects, Rawls observed the political tumult of his era from a remove as great as Kant's, in provincial Königsberg, despite the fact that the Cambridge, Massachusetts, of Rawls's time was intimately connected with the American seat of power in Washington, DC. Rawls, though very much the professional academic, admired John Stuart Mill's pursuit of a vocation as "an educator of enlightened and advanced opinion," harboring "no wish to become a political figure or a man of party" (LHPP 251). But, apart from adding his name to a scattered handful of petitions over the course of his years, Rawls showed no inclination whatever to play the role of public intellectual.

Given all this, it is less surprising that the development of Rawls's thinking was not directly responsive to any particular political controversies that emerged or culminated during his adulthood. The sole exceptions in his published *opera* are his few pages of criticism of the U.S. Supreme Court's campaign finance jurisprudence, particularly the 1976 decision in the *Buckley v. Valeo* case, and his analysis of the question of justifying civil disobedience – a hot topic in the 1960s – which mainly preceded the appearance of *A Theory of Justice*, in 1971.

The two most conspicuous, and commented-upon, changes in his 1971 view were an extension of the account of justice to the international case and an overhaul of the account of the stability of a well-ordered society. The former change was responsive to the widely voiced criticism that his focus on the question of justice of a self-contained, economically advanced society left out of account the more pressing issues of the day: those of global justice and the inequalities of life prospects that flow from the happenstance that some are born in rich countries and others in poor ones. The publication in 1999 of *The Law of Peoples* fulfilled the promise in *A Theory of Justice* to extend the theory to the international case. As it turned out, the extension drew even greater criticism than the omission, much of it exasperatedly negative rather than constructive; but that is not the subject of the present book. Here, the focus is on the theory of justice for an idealized, self-contained society, and on the roads to achieving such a society.

The latter of the two prominent changes was the adjustment in the account of stability, and the coordinate emphasis on a political rather than a comprehensive form of liberal theory. Unlike the first change, this one seems to have been almost entirely self-motivated. Rawls was scrupulous about crediting critics for forcing him to clarify or adjust his positions, but in the Introduction to *Political Liberalism* no one is credited with having pointed out to him the "serious problem internal to justice as fairness, namely . . . the fact that the account of stability in part III of *Theory* is not consistent with the view as a whole" (PL xvii–xviii). (Rawls does thank his Harvard colleague Burton Dreben for helping him find his way along this path, once he had taken it.) It is remarkable that this major inconsistency had not been

pointed out in the already extensive critical literature. Perhaps this is a clue that Rawls's concern about a deep inconsistency was merely valetudinarian; or, more charitably, it could be seen as further evidence of the depth of his insight. Rawls's continuing focus on stability also led him, in *Justice as Fairness: A Restatement*, to reconfigure the argument for the two principles of justice as fairness, in ways that have not yet been sufficiently remarked.

The subject of this book is a related, equally significant development in Rawls's thinking. This is his express insistence that *welfare-state capitalism* cannot satisfy the two principles of justice as fairness. The significance of this change is underscored by the fact that numerous commentators had understood *A Theory of Justice* to have been designed to serve as a defense of welfare-state capitalism. Rawls's Harvard colleague Hilary Putnam recalled that "the publication of *A Theory of Justice* . . . coincided with enormously important debates in American public life about the rightness or wrongness of the welfare state" (Putnam 1997, 189). Americans had by that time become accustomed to thinking of their nation as a welfare state, and work such as Daniel Patrick Moynihan's *The Negro Family: The Case for National Action* (popularly known as "The Moynihan Report" [1965]) was widely taken as a solemnly irreverent challenge to the received wisdom of remaining a welfare state – even to the morality of remaining so. The relatively minor portion of *A Theory of Justice* devoted to certain questions of "non-ideal theory" contains passages that strongly suggest that Rawls considered the United States circa 1971 to be a "reasonably just" or "nearly just" state (*TJ* 308, 309). So, for these and many similar reasons, it was easy to view the work as a "transcendental deduction" of welfare-state capitalism – that is, as a defense of the basic justice of a society that allows a normally only lightly regulated market to determine prices, wages, and capital investment, and which makes separate provision to attend to those whose participation in the labor market is impossible or insufficient to meet their needs.

In 1988, Richard Krouse and Michael McPherson (1988, 79 n.1) compiled an impressive list of scholarly commentators who understood Rawls this way: Robert Paul Wolff, Brian Barry, Allen Buchanan, Barry Clark, Herbert Gintis, Norman Daniels, Amy Gutmann, Carole Pateman, Alan Ryan, and David Schweikert. Krouse and McPherson cite but one writer, Arthur DiQuattro (1983), as having perceived that *justice as fairness*, the theory Rawls advanced in *A Theory of Justice*, is unfriendly to capitalism. DiQuattro himself had already listed the names of C. B. Macpherson, Benjamin Barber, Robert Amdur, and Kai Nielsen on the roll of non-perceivers of Rawls's antipathy to capitalism. Austrian economist and free-market champion Friedrich Hayek was satisfied that *A Theory of Justice* is not properly interpreted "as lending support to socialist demands" (1984, 113). Even the sympathetic Marxist Rodney Peffer wrote that "Rawls *presumes* that a democratic form of welfare state will best conform to . . . the requirements of social justice" (2014 [1990], 378; reordered, but emphasis in the original).

DiQuattro rightly pointed out that the word “capitalism” does not occur at all in *A Theory of Justice*. But the expressions “private-property economy,” “private-property society,” “private-property regime,” and “private-property democracy” are used, and are used in explicit contrast to “socialism” (TJ 235–236, 239, 242). Rawls is quite clear that his theory of justice – I will call it “justice as fairness,” as was Rawls’s practice, or “the two principles”³ – regards economic inequality as an inescapable fact of life. Justice as fairness is designed to make extensive use of a labor market, and it legitimates wage and salary differentials as incentives that are necessary to achieve economic efficiency. Rawls took the trouble to show how “a properly organized democratic state that allows private ownership of capital and natural resources” could “fit the two principles of justice” (TJ 243–251). In a properly designed private-property economy, with “background justice” assured, and so long as there is also a “social minimum” safety net, “the distribution of wealth that results is just whatever it is” (TJ 249). And Rawls ruled out by all but name the Communist or “command” socialist economies that were familiar, at least in broad outline, to contemporary readers, on the ground that they infringe the “important liberty of free choice of occupation” (TJ 242), not to mention the equal political liberties. The many commentators who took Rawls as a friend of welfare-state capitalism were not guilty of any crude misreading of the text.⁴ It is not straining to read Rawls’s social minimum and difference principle as intended to serve as sufficient correctives of the inequalities inevitably generated by a just market system of private ownership.⁵ Classifying the theory as a defense of welfare-state capitalism was an easy mistake to make, if indeed that is what it was.

³ Throughout *A Theory of Justice*, Rawls also refers to the theory developed therein as “the contract theory.” I will not do so. Rawls later emphasized that justice as fairness is not unique in being a theory of justice supported by contractualist methods of theory-building, and he stopped referring to the theory of *Theory* as “the contract theory.”

⁴ Crude misreadings tended instead to lead the opposite way. Sociologist Daniel Bell, for example, saw *A Theory of Justice* as “the most comprehensive effort in modern philosophy to justify a socialistic ethic” (1972, 72). But note Bell’s reading of the difference principle:

We have here a fundamental rationale for a major shift in values; instead of the principle “from each according to his ability, to each according to his ability,” we have the principle “from each according to his ability, to each according to his need.” And the justification for need is fairness to those who are disadvantaged for reasons beyond their control.

(1972, 57)

In fairness to Bell, I note that he was writing for an educated but not a specialist readership, and was attempting to weave Rawls’s then very recent big book into a broad account of the intellectual history behind an affirmative-action debate that was then also new. Conflating Rawls’s difference principle with Marx’s prescription for the “higher phase of communist society” (1978 [1875], 531) would not ordinarily be excusable.

⁵ Political philosopher Michael Otsuka takes Rawls as an exemplar of “liberal egalitarianism” i.e., as avowing a stringent right of control over one’s own mind and body but denying any “very stringent right to all of the income that one can gain . . . on one’s own or through unregulated and untaxed exchanges” (2003, 15 n.17). “Liberal egalitarian capitalist” is not an oxymoron.

Rawls accepted responsibility for this misunderstanding. In the 1987 preface to the French translation of *A Theory of Justice*, Rawls confessed that he would now “distinguish more sharply the idea of a *property-owning democracy* . . . from the idea of a welfare state . . . [which is] quite different” (CP 419; emphasis added). He warned that a welfare state “*may allow* large and inheritable inequalities of wealth incompatible with the fair value of political liberties . . . as well as large disparities of income that violate the difference principle” (CP 419; emphasis added). This passage was repeated in the preface to the revised edition, published in 1999 (TJ xv). The “may allow” could be read as merely cautionary, or – as his colleague and “tutor” Burton Dreben might point out – it could also be read as intimating in a cramped, muffled way that capitalism, in both its *laissez-faire* and its welfare-state guises, tends inevitably toward injustice.

Burt was doing what he could to make me be clear, to write forcefully and sharply, to be less guarded and muffled, a term he often used. He wanted me to find my own “voice,” as we sometimes say. He would often comment by name on other people who were extraordinarily bright and knowledgeable but failed to express themselves clearly and with vigor. Their style was muffled and cramped, somehow they held back.

(Rawls 2000a, 426)

It is fairly certain that Dreben had no opportunity (if he had the desire) to press Rawls to be more vigorous in this particular connection.

Finally, only shortly before his death in 2002, Rawls for the first time declared in print, in *Justice as Fairness: A Restatement* (2001), that welfare-state capitalism *could not* realize justice as fairness even in the best of circumstances (JF 135–138). It is a “serious fault” (JF 139 n.5) of *Theory* that it does not emphasize that welfare-state capitalism is not among the private-property regimes that might satisfy the two principles. Put bluntly, capitalism in all its guises is unjust. But the *Restatement* did not excite the curiosity of the educated public, or even of specialists, to the same degree as *A Theory of Justice* had, nor even as much as had the *Restatement*’s predecessors, *Political Liberalism* (1st ed. 1993) and *The Law of Peoples* (1999). The often astute Cambridge political philosopher Raymond Geuss, for example, wrote that *Law of Peoples* is Rawls’s “last systematic work” (2005, 33).

Possibly this diminution of interest was symptomatic of a general “Rawls fatigue” that was accumulating – along with the enormous secondary literature – over the thirty years since 1971.⁶ Jeremy Waldron, the former Chichele Professor in Social and Political Theory at Oxford, wants “to encourage young political theorists to understand that there is life beyond Rawls” (2016, ix). Waldron means no disrespect, but his advice spells relief for theorists of all ages: it was hard even for sympathetic

⁶ This literature’s growth owed in part to debates premised in part on other misreadings of Rawls, as philosopher Elizabeth Anderson, his former student, recounts (Anderson 1999; Anderson forthcoming; see also Scheffler 2003).

readers to keep up with the spate of Rawls titles issued by Harvard shortly before his death, including the revised edition of *A Theory of Justice*, *Collected Papers*, and *Law of Peoples* of 1999. Set aside first editions, and the *earliest* of Rawls's books is the expanded edition of *Political Liberalism* that came out from Columbia in paperback in 1996, fully a quarter-century after the first edition of *Theory*.

Rawls did not exploit the obvious moment to advertise his wholesale rejection of capitalism. His confession of “the more serious faults” of *Theory*, tendered in the preface to the *Restatement*, does not even mention the “serious fault” in *Theory* that enabled his misclassification as an apologist for welfare-state capitalism. Nor does the preface cite his rejection of welfare-state capitalism as one of the “main changes” (JF xvi) the *Restatement* performs. It is therefore not at all surprising, much less, inexplicable, that the reputation of “John Rawls, apologist for welfare-state capitalism” persists.

And persist it does. Consider these two recent, dismissive references to Rawls, the apologist for late capitalism. According to the Swiss political philosopher Christoph Henning (2014, 461),

The moment the individuals in the original position begin to make concrete decisions, the institutions they create resemble those of the USA down to the smallest detail. *A Theory of Justice* needs therefore to be read as a transcendental deduction of the USA [circa 1971].

“Bourgeois practical philosophy” is the category to which the “Philosophical Gourmet,” Brian Leiter, assigns Rawls. Although he approves Rawls's focus on the basic structure of society rather than on individual choices, Leiter assumes that “the basic structure does not include capitalist relations of production” (2015, 32 n.15), and concludes that

while Rawls ... endorsed intuitions that had implications for basic social and economic policy in capitalist societies, his theory was neither presented nor understood as threatening capitalist relations of production, a fact surely central to any explanation of how it could become so influential in capitalist democracies, at least in the universities.

(Leiter 2015, 31)

Not only is Rawls an apologist for capitalism, Leiter alleges, Rawls also owes his influence as a philosopher to his being an apologist for capitalism. My view differs. I am confident that Rawls's influence will not suffer when his theory's profound hostility to capitalism is appreciated. Rawls's theory not only does not legitimate the institutional status quo in the United States, as it was in 1971 or as it is today; it condemns it as fundamentally unjust and demands a radical reconstitution.

Readers whose primary acquaintance with Rawls comes from *Theory* and *Political Liberalism* may not be aware that the *Restatement* also represents an unheralded and complex reformulation of the argument for the two principles of justice as fairness. In the introduction to the paperback edition of *Political Liberalism*, Rawls conceded that

certain alternatives to the two principles of justice as fairness – for example, liberal conceptions that reject the difference principle – can count as reasonable liberal political conceptions (*PL* xlvi–xlix; cf. Metz 2002, 619). This could have been misunderstood as a sign of surrender rather than the mere bracketing of the issue. In the *Restatement*, Rawls returns to the affirmative. He responds to criticism by economists John Harsanyi, Kenneth Arrow, and others by making more careful and sparing use of the “maximin” criterion of decision under uncertainty. In its place – taking up the slack, as it were – Rawls relies much more than he does in *Theory* on “such ideas as publicity and reciprocity” (*JF* xvii). It will emerge that stability, publicity, and reciprocity also figure centrally in the argument for socialism that Rawls adverts to but never detailed.

To recapitulate very, very briefly, and solely as a reminder of what many readers will rightly feel they already know by heart: *A Theory of Justice* presented a “working up” of ideas already pronounced in the political culture of modern constitutional democracies, enabling an ideally rational and reasonable chooser to derive two principles of justice from the “original position,” the device for assuring that the choice of principles not be warped by bias. Those principles, as initially stated, are:

First: each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all.

(*TJ* 53)

The two principles operate in what he calls a lexical order: satisfying the first principle has an absolute priority over the second, such that first-principle liberties – although to be adjusted to fit together – are not to be sacrificed to further second-principle values. Rawls made refinements and adjustments in the two principles, in ways that will be relevant later in the book.

In the *Restatement*, Rawls gives greater prominence to the division of the argument for the two principles into two parts (*JF* 87–89; 180–181; cf. *TJ* 124–125, 464–465). One reason for the division is to simplify the task facing the choosers in the original position, by freeing them from concern with what he calls the “special psychologies” of the persons the choosers represent. In Part One, the parties choose principles without knowing that the persons whom they represent are prone to “be envious or spiteful, or to have a will to dominate or a tendency to be submissive, or to be peculiarly averse to uncertainty and risk” (*JF* 180). Part Two of the argument takes the two principles as provisionally established, and the parties return to the problem of stability. The parties

now consider the special psychologies by checking whether those who grow up under just institutions (as the principles adopted specify them) will develop a sufficiently firm sense of justice with respect to those attitudes and inclinations.

(*JF* 184)

That is to say, the choosers must ask whether citizens in a well-ordered society will acquire a sufficiently strong and effective sense of justice so that they normally comply with just arrangements and are not moved to act otherwise, say, by social envy and spite, or by a will to dominate or a tendency to submit. (JF 181)

If the choosers can satisfy themselves that the principles of justice support a sense of justice sufficient to resist the “destabilizing special attitudes” (JF 181), then “the outcome of the first part of the argument is confirmed and the argument for the two principles is complete” (JF 181).

After the derivation of the two principles of justice, but before introducing the special psychologies, Part One of the argument outlines a “four-stage sequence” that proceeds from the original position, to a constitutional stage at which the parties “decide upon the justice of political forms and . . . design a system for the constitutional powers of government and the basic rights of citizens” (TJ 172), to a legislative stage, where majority rule is introduced, to a final, administrative *cum* judicial stage. In the *Restatement*, Rawls introduced a new task at the second, constitutional stage. For the first time, he considered five “ideal-types” of regime as candidate institutional forms for realizing justice as fairness:

laissez-faire capitalism,
 welfare-state capitalism,
 one-party “state” socialism,
 property-owning democracy, and
 liberal (democratic) socialism (JF 136).

Part One of the original position procedure is not complete until this is undertaken:

It is . . . important to trace out, if only in a rough and ready way, the institutional content of the two principles of justice. We need to do this before we can endorse these principles, *even provisionally*.

(JF 136; emphasis added)

By “even provisionally” Rawls is acknowledging that checking the “institutional content” is necessary in both Part One of the original position procedure, where the citizens’ special psychologies are not known to the parties, and in Part Two, where the parties then take the special psychologies into account, and focus again on the question of stability. The “realist” critics, dissatisfied with Rawls’s focus on “ideal theory,” had insisted that political philosophy ought to stress “the evaluation and comparison of institutions and regime types, not only principles” (Galston 2010, 408); what Rawls offered could be taken as a response to this challenge, but Rawls presented it as already required in ideal theory by the procedure of reflective equilibrium.

After setting out the list of five ideal regime types, Rawls itemized four questions that arise with respect to any of them. One: whether it would be “right and just” on the assumption that it could be “workably maintained.” Two: whether its “institutions can be effectively designed to realize its declared aims.” Three: “whether citizens, in view of their likely interests and ends as shaped by the regime’s basic structure, can be relied on to comply” with its institutions and rules – a question that includes “the problem of corruption.” Finally, four: “there is the question of competence: whether the tasks assigned to offices and positions would prove simply too difficult for those likely to hold them” (*JF* 136). Rawls sets aside all but the first question. The three further questions will have to be dealt with in both parts of the original position procedure. Rawls is thus, in effect, limiting his “illustrative and highly tentative” (*JF* 136) survey of the five regime types to scrutiny under Part One on the assumption that each can “be effectively and workably maintained,” which is an assumption that still must be scrutinized under Part One and, again under Part Two.

Rawls asks: “When a regime works in accordance with its ideal institutional description, which of the five regimes satisfy the two principles of justice?” (*JF* 137). By making the further assumption that “if a regime does not try to realize certain political values, it will not in fact do so” (*JF* 137), Rawls quickly dismisses one-party socialism and laissez-faire capitalism. One-party socialism does not protect equal political liberty or freedom of occupation, and laissez-faire capitalism does not aim to secure the fair value of political liberty and fair equality of opportunity. No “invisible hand” possibility is entertained.

This much was easily foreseeable; what is new is that Rawls went on to reject welfare-state capitalism. Welfare-state capitalism cannot realize the two principles of justice as fairness because it does not express reciprocity between citizens as free equals, it is not serious about fair equality of opportunity, and it does not guarantee, or even try to guarantee, the fair value of the political liberties. The *Restatement* makes more prominent the importance Rawls had placed, in *Theory*, on the fair value of political liberty as a prior, first-principle guarantee of reciprocity between citizens as free equals (cf. Wall 2006).

So, now having dismissed welfare-state capitalism, command-economy socialism, and laissez-faire capitalism, “this leaves . . . property-owning democracy and liberal socialism: their ideal descriptions include arrangements designed to satisfy the two principles of justice” (*JF* 138). Rawls states that he does not believe that justice as fairness can decide at this point between the two remaining ideal-types of regime: property-owning democracy and liberal socialism.

When a practical decision is to be made between property-owning democracy and a liberal socialist regime, we look to a society’s *historical circumstances*, to its *traditions of political thought and practice*, and *much else*. Justice as fairness does not decide between these regimes but tries to set out *guidelines* for how the decision can reasonably be approached.

(*JF* 139; emphases added)

This is disappointing, insofar as Rawls's task is "to see whether we can resolve the impasse in our recent political history; namely, that there is no agreement on the way basic social institutions should be arranged if they are to conform to the freedom and equality of citizens as persons" (*PL* 300).⁷ At least since 1848, the most intractable impasse in political thought, by far, has concerned the ownership of socially significant assets – the means of production – and the competing claims made for state and for private proprietorship. The central thesis of this book is that, despite Rawls's seeming disclaimers, he was aware that his ideal theory does in fact contain sufficient resources to resolve this central impasse, and to resolve it in favor of liberal democratic socialism. This becomes evident when the question of stability is addressed, and pursued under both Part One and Part Two of the original position procedure.

The "historical circumstances" and "traditions of thought and practice" referred to in the block quotation above are already among the facts revealed to the parties at the constitutional stage. The "much else" is presumably a reminder of the "highly controversial" and "intricate" questions about "public funding of elections and political campaigns, different kinds of property ownership and taxation" (*JF* 136) that he did not want to take up either. The "guidelines" Rawls refers to are not further specified, but presumably they are the "guidelines of public reason" essential to a political conception of justice, which are adopted in the original position as companions to the two principles (*PL* 223–225). Rawls's desire is that these guidelines, taken together with the substantive content of justice as fairness, be "complete":

This means that the values specified by [the political] conception [of justice] can be suitably balanced or combined, or otherwise united . . . so that those values alone give a reasonable public answer to all, or nearly all, questions involving the *constitutional essentials* and *matters of basic justice*.

(*PL* 225; emphasis added)

"Constitutional essentials" are matters not included in the two principles, but they "concern questions about what political rights and liberties, say, may reasonably be included in a written constitution, when assuming the constitution may be interpreted by a supreme court, or some similar body." By contrast, what Rawls terms "matters of basic justice" are subjects that "relate to the basic structure of society and

⁷ Rawls said even less on the subject in the 1989 *Guided Tour*, breaking off after merely noting that property-owning democracy and liberal socialism were alone left standing (*GT* 112). It is therefore a fair inference that Rawls had either given no further thought to the question of deciding between these two surviving regime types, or was not yet ready to put forward, in written form, what he thought. Rawls is said to have written and circulated the *Guided Tour*, as a photocopy, for the benefit of Harvard Law students. With few changes, it was ultimately published as the *Restatement*.