

Introduction

In his essay “Two Concepts of Liberalism,” William Galston distinguishes between two varieties of liberal theory.¹ The first—Enlightenment liberalism—stresses the development and exercise of our capacity for *autonomy*, understood as “individual self-direction” and entailing a “sustained rational examination of self, others, and social practices”; this is the liberalism of not only Kant and Mill but also a number of contemporary thinkers, including Don Herzog, Stephen Macedo, Jeremy Waldron, and the preeminent Kantians (Barbara Herman, Christine Korsgaard, Onora O’Neill, Allen Wood, etc.).² The second—Reformation liberalism—emphasizes *diversity* and the toleration that encourages it, where diversity is understood simply as “differences among individuals and groups over such matters as the nature of a good life, sources of moral authority, reason versus faith, and the like”; this is the liberalism of not only Madison and Isaiah Berlin but also contemporary thinkers such as Galston himself, Charles Larmore, and Donald Moon.³ These two varieties of liberal theory are often mutually supporting—as Galston puts it, “the exercise of autonomy yields diversity, while the fact of diversity protects and nourishes autonomy”—but in a surprising number of cases they conflict, whether over the accommodation of group difference, the design of civic education, or the promotion of liberal values internationally.⁴ In fact, much of the so-called “liberalism/multiculturalism debate” is an intramural affair, pitting Enlightenment

¹ Galston 1995. His distinction between “Enlightenment” and “Reformation” liberalisms was anticipated by Charles Larmore’s distinction between “Kantian” and “modus vivendi” liberalisms and Donald Moon’s distinction between “traditional” and “political” liberalisms, respectively; see Larmore 1987 and Moon 1993.

² Galston 1995, 521, 523, 525. He identifies Herzog, Macedo, and Waldron as Enlightenment liberals.

³ Ibid., 521, 525-7. He identifies himself, Madison, and Berlin as Reformation liberals, at least implicitly. Locke is harder to categorize. The *Letter Concerning Toleration* has both Enlightenment and Reformation components: some of its arguments focus on the idea that only a “free faith” can have any worth in the eyes of God, while others place emphasis on the peace and security that will follow from toleration of diverse sects—see Locke 1990, 19, 65, 71.

⁴ Galston 1995, 521. Regarding the third case, see Mehta 1999 on Mill, Burke, and British colonialism.

and Reformation liberals against one another.⁵

One might reasonably ask where John Rawls, arguably the greatest political philosopher of the twentieth century, would fall in this debate. He certainly had many Enlightenment-liberal credentials: he taught several famous Kantians (e.g., Herman, Korsgaard, and O'Neill), lectured on Kant extensively, and characterized his *magnum opus*, *A Theory of Justice* (1971), as “highly Kantian in nature.”⁶ By the same token, though, Rawls’s later work *Political Liberalism* (1993) “applies the principle of toleration to philosophy itself,” thus taking a diversity-based approach that has been a major influence on such Reformation liberals as Galston, Larmore, and Moon.⁷ We might therefore understand Rawls’s intellectual trajectory as the opposite of the historical one: it begins with the Enlightenment and ends by circling back to the Reformation.

This depiction of his trajectory is far too crude, however. *Political Liberalism* may be a Reformation-liberal text, but is *Theory* really an Enlightenment-liberal one—or, more precisely, is it a Kantian-liberal one? Many scholars have called Rawls’s Kantian credentials into question, including but not limited to Kerstin Budde, Otfried Höffe, Oliver Johnson, Larry Krasnoff, and Andrew Levine.⁸ Other scholars (e.g., Larmore) have discerned certain justificatory ambiguities in *Theory*, such as the commingling of Enlightenment-liberal and Reformation-liberal elements.⁹ Most importantly, Rawls himself saw a strong continuity between the arguments of *Theory* and *Political Liberalism*, suggesting that the Kantianism of the former work may have been oversold,

⁵ See, for example, Laden and Owen 2007, as well as the discussion in Kymlicka 2002, Chapter 8.

⁶ TJ xviii. The Kant lectures take up approximately half of LHMP. A number of Rawls’s students, including those listed above, contributed to an edited volume that took its inspiration from Rawls’s approach to teaching the history of moral and political philosophy—see Reath, Herman, and Korsgaard 1997.

⁷ PL 10. Galston, however, denies that the later Rawls is a Reformation liberal, arguing that he “attempts to give due weight to our deepest differences [but] ultimately fails to take those differences seriously enough” (1995, 518-21).

⁸ See Budde 2007, Höffe 1984, Johnson 1974, Krasnoff 1999, and Levine 1974.

⁹ Larmore 1987, 125: “*Theory of Justice* harbors, side by side, the Kantian and modus vivendi approaches. Rawls’s later writings, and particularly his Dewey Lectures [i.e., KCMT], have put the second approach in the center where it belongs.”

not only by himself but by others as well.¹⁰

I will therefore begin in Chapter 1 by showing just how Kantian Rawls was during his most Kantian period—roughly, from *Theory of Justice* to his “Kantian Constructivism in Moral Theory” (1980) and “Social Unity and Primary Goods” (1982).¹¹ I demonstrate here that Rawls’s theory is even more Kantian in this period than has generally been recognized: from his Kantian conceptions of person and society to his construction procedure (including the formal constraints of the concept of right, the veil of ignorance, and the thin theory of the good) and on through the principles, institutions, and psychology of justice that this procedure generates, his insights track those of Kant nearly one for one. Alternative readings of *Theory* and other works of this period are possible—as Rawls himself argues and as I will show in Chapter 7—but their essentials are profoundly and almost unremittingly Kantian.

The interpretive work of Chapter 1 provides the essential backdrop for the reconstructive task of Part 2 (Chapters 2-6), which shows just how dependent Rawls’s arguments for the most distinctive features of justice as fairness—namely, the lexical priorities of right, political liberty, civil liberty, and fair equality of opportunity plus the difference principle—are upon his extreme and controversial Kantian conception of the person. Its very extremity and controversiality will only become clear, however, in the process of (re)constructing his arguments for these features: through a procedure of “backwards engineering,” I will show that any conception of the person that is capable of grounding the arguments for these features must be one that is itself grounded

¹⁰ JFPM 388-9. Rawls explicitly “put aside the question whether the text of *A Theory of Justice* supports different readings from the one I sketch here” (388). One of the most important tasks of my book will be to provide just such an alternative reading. Among those who seconded Rawls’s self-described Kantianism were Darwall 1976 and 1980, Davidson 1985, and Guyer 2000, 262-86.

¹¹ Samuel Freeman identifies Kantian constructivism as a “transition stage” in Rawls’s thought (CP xi). Stephen Darwall (1980) is the only other scholar to have carried out a *comprehensive* examination of Rawls’s Kantianism, but he did not have access to KCMT when he wrote his essay. My Chapter 1 is therefore the first comprehensive examination of Rawls’s Kantianism in light of his Kantian-constructivist writings, which do much to elucidate the nature (and limits) of his Kantianism.

in Kant's model of finite rational agency, properly elaborated. These reconstructions of Rawls's arguments are required because either (1) they are incomplete (as with the priority of liberty and the difference principle) or (2) they are basically missing (as with the priority of fair equality of opportunity). Moreover, the principles of justice that these arguments sustain play such a central, consistent role in the various incarnations of Rawls's theory—from *Theory* (1971) and *Political Liberalism* (1993) to *Justice as Fairness: A Restatement* (2001)—that flaws in their justification imperil his evolving political project.¹² This reconstructive task occupies the heart of my book, supported by the prior interpretive work of Chapter 1, which not only offers materials for this task but also assures us that a Kantian reconstruction does no violence to Rawls's texts.

I begin this task in Chapter 2 by presenting a more detailed and comprehensive Kantian conception of the person than Rawls uses in his own works, but one that is required to justify his theory's most distinctive features. This conception of the person, based upon Kant's own model of finite rational agency, is a hierarchy of Kantian conceptions of autonomy: in descending order, they are Kantian moral autonomy, Kantian personal autonomy, and Kantian self-realization. The constituent conceptions of this hierarchy offer the most compelling interpretations of their parent concepts and can be reached along both deductive and inductive routes. The first route utilizes a procedure analogous to Rawls's four-stage sequence to derive the lower conceptions from higher ones along with their associated plans and rules, while the second constructs an ideal cognitive-developmental psychology, an epicyclic system using an iterative model of agency to explain the emergence of higher conceptions from lower ones.¹³ This hierarchy of Kantian conceptions will be used in later chapters to ground a parallel hierarchy of lexical priorities, with the priorities of

¹² For example, Rawls devotes a huge amount of text to defending the priority of liberty and the difference principle in *Political Liberalism* and *Justice as Fairness: A Restatement*, respectively; see PL 289-371 and JF Parts II and III.

¹³ Rawls's four-stage sequence is described at TJ §31. This ideal cognitive-developmental psychology loosely tracks Rawls's own discussion of psychological "stage theories," such as those of Piaget and Kohlberg, in TJ Chapter 8.

right and political liberty at the top, the priority of civil liberty in the middle, and the priority of fair equality of opportunity (or FEO) at the bottom. This rich system of relationships is depicted graphically in Figure 1 below. Notice that all priorities as well as their grounding conceptions of autonomy can be traced back to a Kantian conception of moral autonomy—a result anticipated by Kant in his *Groundwork*, as I shall argue.¹⁴

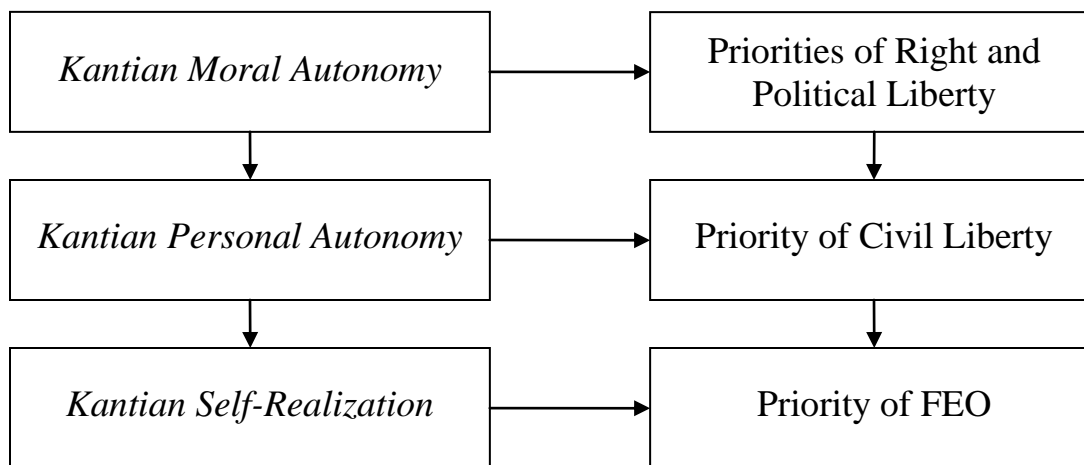


Figure 1: Parallel Autonomy and Priority Hierarchies

Chapters 3 through 6 then review the various arguments Rawls makes for the distinctive features of his theory and show that only those arguments securely grounded on the hierarchical conception of persons presented in Chapter 2 offer genuine promise. Chapter 3 does this for the priorities of right and political liberty, arguing that they are founded on a Kantian conception of moral autonomy and also that (contrary to Rawls’s assertion in his famous “Reply to Habermas”) political liberty takes priority over civil liberty in his theory, giving it a somewhat civic-humanist character.¹⁵ Chapter 4 reveals the grave inadequacy of two of his three arguments for the priority of (civil) liberty and shows how the third and most promising one, which grounds this priority on

¹⁴ See GMM 4:415-7.

¹⁵ PL 206, 413.

a Kantian conception of personal autonomy, can be bolstered. Chapter 5 offers for the first time a defense of the priority of fair equality of opportunity—which is entirely (and peculiarly) missing from the corpus of Rawls’s texts—a defense that is constructed with existing resources in *Theory* (including the Aristotelian Principle and the concept of Humboldtian social union) and grounded on a Kantian conception of self-realization through work, which has Marxist undertones. Finally, Chapter 6 explicates and gives additional support for his notorious difference principle. I contend here that his latter-day defenses of the difference principle (especially those present in *Justice as Fairness: A Restatement*) are inadequate and that the most compelling case for it can be made by means of a reconstructed version of an earlier defense—*viz.*, the formal one elaborated in *Theory* §26.¹⁶ This reconstructed defense relies, though, on a whole series of Kantian assumptions about asceticism, the unavoidable ends of duty and finitude, and the essential nature of persons. As one moves through these five reconstructive chapters, the extent to which justice as fairness depends upon specifically Kantian presuppositions becomes increasingly apparent, strongly reinforcing the conclusions of Chapter 1.¹⁷

In Part 3 of the book, I turn to “reflective equilibrium,” Rawls’s innovative technique of moral justification, and to its role in vindicating his Kantian conception of the person. In Chapter 7, I describe this technique, developed in both *Theory* and “The Independence of Moral Theory” (1975), and point to its ambiguous treatment of moral objectivity. This ambiguity is reflected in Rawls’s evasive answer to a central question: how can we bring about a coincidence of reflective judgments among moral agents, which he maintains is “a necessary condition for objective moral

¹⁶ See especially TJ 132-5.

¹⁷ Cf. S. Freeman 2007a, 183, where he maintains that “this deep Kantian argument plays no central role in solving the problems, dealt with in parts I and II of *A Theory of Justice*, of (1) eliciting the reasonable principles of a just constitution, and (2) deciding the institutions that satisfy them.”

truths”?¹⁸ One answer suggested by Rawls—namely, a reliance on “self-evident first principles,” like Kant’s practical postulate of freedom—is rejected during the 1980’s in favor of another: by way of pre-existing (near) consensus on considered convictions of justice found in “a democratic society under modern conditions.”¹⁹ I enumerate Rawls’s powerful reasons for rejecting the first answer and then describe in great detail the mature theory that he eventually developed from the second answer in *Political Liberalism*.

In Chapter 8, I contend that the Reformation-liberal answer given in *Political Liberalism* to the above question—roughly, that justice as fairness can serve as the focus of an overlapping consensus of the reasonable comprehensive doctrines present in modern liberal democracies—is impoverished in at least two senses. First and more narrowly, no comprehensive doctrine but the Kantian one is capable of endorsing the strong Kantian conception of the person that underwrites justice as fairness. Thus, no overlapping consensus on justice as fairness is possible; the best that can be hoped for is an overlapping consensus on a *class* of liberal political conceptions of justice, with justice as fairness as just one competitor conception among others, its centrality determined through political competition and the strength of the supporting socio-economic interests. Second and more broadly, even if we assumed justice as fairness could act as the focus of an overlapping consensus, the system of justification involved would offer us little moral guidance, whether in a domestic or an international context, regarding the appropriate width and content of overlapping consensus: because political liberalism has no *independent* criterion of reasonableness, it cannot refuse extension of the scope of toleration to include illiberal, even indecent groups and nations. Moreover, its agnosticism regarding the width and content of overlapping consensus, which is a form of cultural relativism, would lead to a dramatically diminished role for political philosophy,

¹⁸ IMT 290.

¹⁹ IMT 289; KCMT 305-6.

effectively turning it into a handmaid of the social sciences. These are sharply revisionist claims, of course, but I believe they are borne out by the chapter's arguments, which as a whole suggest the profound poverty of political liberalism and Reformation liberalism more generally.

Finally, the Conclusion offers an alternative way to answer the question of justification, showing that a Kantian conception of the person must, unsurprisingly, be grounded in a Kantian way: on a practical postulate of freedom as a necessary presupposition of finite rational agency. I argue here, however, that this defense need not rely upon Kant's transcendental idealism, i.e., we can dispense with metaphysics in grounding Kantian liberalism—even the “thin” metaphysics of Kant—rendering the postulate potentially ratifiable by a wide variety of persons as well as belief systems.²⁰ This justificatory approach is Rawls's “road not taken,” and by following it ourselves, we can turn justice as fairness into an authentically comprehensive and universalistic liberalism, thereby fulfilling the implicit promise of *Theory* and helping secure its place within the canon.²¹

Even if this reconstruction of Rawls's Kantian liberalism is found unconvincing by some persons, it can still be seen by them as a worthy companion to other cosmopolitan Enlightenment liberalisms—e.g., Millian plural-perfectionism, Benthamite or Sidgwickian liberal utilitarianism, and Lockean religious liberalism—in the fight against illiberal principles and institutions around the globe, including theocratic, secular authoritarian, and totalitarian regimes and their sustaining ideologies. While Reformation liberalism is virtually impotent in such contexts, too unsure of its own relevance in illiberal societies, universalistic Enlightenment liberalisms are not hobbled by such doubts. Their very diversity is a source of strength, in fact, because different Enlightenment

²⁰ Rawls himself claims to offer a detranscendentalized Kant, one contained “within the framework of an empirical theory” (TJ 226-7). While I share Rawls's aspirations, I believe that he dispenses with too much of Kant's practical philosophy in the process: a genuinely *Kantian* liberalism must be based upon a practical postulate of freedom, as I will argue in the conclusion; without this presupposition of finite rational agency, justice as fairness will simply be one more variety of heteronomous liberalism, retaining the form but not the substance of an autonomous theory.

²¹ Frost 1969, 105.

liberalisms are likely to appeal to different individuals, groups, and societies. As I contend in the conclusion, these liberalisms offer us the vision of a liberal world order (“a republicanism of all states, together and separately,” as Kant put it) and a mode of justification addressed to all men and women as human beings, not as members of various religious, racial, and national groups.²² Their optimistic cosmopolitanism makes them worthy of our allegiance and—if the arguments of this book are sound—makes the Kantian liberalism of a reconstructed Rawls the most worthy of them all.

²² MM 6:354.