

Why Political Liberalism?

On John Rawls's Political Turn

Paul Weithman



Why Political Liberalism?

OXFORD POLITICAL PHILOSOPHY

GENERAL EDITOR: SAMUEL FREEMAN, UNIVERSITY OF PENNSYLVANIA

Oxford Political Philosophy publishes books on theoretical and applied political philosophy within the Anglo-American tradition. The series welcomes submissions on social, political, and global justice, individual rights, democracy, liberalism, socialism, and constitutionalism.

N. Scott Arnold

Imposing Values: An Essay on Liberalism and Regulation

Peter de Marneffe

Liberalism and Prostitution

Debra Satz

Why Some Things Should Not Be for Sale
The Moral Limits of Markets

William J. Talbott

Human Rights and Human Well-being

Paul Weithman

Why Political Liberalism? On John Rawls's Political Turn

Why Political Liberalism?

On John Rawls's Political Turn

Paul Weithman

OXFORD
UNIVERSITY PRESS
2010

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright (c) 2010 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Weithman, Paul J., 1959-
Why political liberalism? : on John Rawls's political turn / by Paul Weithman.
p. cm.
ISBN 978-0-19-539303-3 (alk. paper)
1. Justice. 2. Liberalism. 3. Political stability. 4. Rawls, John, 1921-2002—Criticism and
interpretation. I. Title.
JC578.W42 2010
320.092—dc22
2009047179
ISBN: 9780195393033

1 3 5 7 9 8 6 4 2
Printed in the United States of America
on acid-free paper

For my teachers

“When fully articulated, any conception of justice expresses a conception of the person, of relations between persons and of the general structure and ends of social cooperation. To accept the principles that represent a conception is at the same time to accept an ideal of the person, and in acting from these principles, we realize such an ideal.”

—John Rawls, “A Kantian Conception of Equality”

Contents

<i>Acknowledgments</i>	xi
List of Tables	xiii
Introduction	3
§1: Overview	4
§2: The Road to Come	8
§3: A Deeper Understanding of Justice as Fairness?	9
§4: Unity, Theodicy, and the Attractions of Liberalism	11
§5: A Final Word to the Reader	14
Chapter I: The <i>Public Basis View</i>	17
§1.1: Initial Statement of the <i>Public Basis View</i>	17
§1.2: The Pivotal Argument	21
§1.3: Imputing the Pivotal Argument?	23
§1.4: The <i>Public Basis View</i> Restated	28
§1.5: Difficulties with the Strong Version	32
§1.6: Difficulties with the Weak Version	36
§1.7: Conclusion	40

Chapter II: Stability and Congruence 42

§II.1: Stability, Inherent and Imposed 43

§II.2: Matching the Right and the Good in Justice as Fairness 51

§II.3: Congruence and Stability 57

§II.4: Congruence and Inherent Stability 65

Chapter III: Ideals and Inconsistency 68

§III.1: An Inconsistency in Justice as Fairness? 70

§III.2: Ideals and Comprehensive Conceptions 72

§III.3: Endorsing on the Basis of Shared Ideals 83

§III.4: Congruence and C_3 88§III.5: C_3 and Inconsistency 96**Chapter IV: The Acquisition of Four Desires 97**

§IV.1: Two Readings of the Aristotelian Principle 99

§IV.2: The Acquisition of Four Desires 103

§IV.3: Four Desires and Thin Reasons 118

Chapter V: Thin Reasons to Be Just 122

§V.1: Setting up the Problem 124

§V.2: The Aristotelian Principle and the Argument for Congruence 127

§V.3: Four Thin Reasons 130

§V.4: Some Questions about the First Three Arguments 141

§V.6: Some Puzzles about the Fourth Argument 146

Chapter VI: The *Argument from Love and Justice* 148

§VI.1: Balances and Temptations 149

§VI.2: Two Questions about Table II.3 153

§VI.3: Conditional Balances and Balance Conditionals 158

§VI.4: The *Argument from Love and Justice* 163

§VI.5: Love's Balance 168

§VI.6: Four Comments on the Argument 176

Chapter VII: Kantian Congruence and the Unified Self 183§VII.1: An Overview of the *Kantian Congruence Argument* 184§VII.2: The Argument from C_4a 188

§VII.3: From the Ostensible Conclusion to Congruence 192

§VII.4: Establishing (5.5') 203

§VII.5: Defending (5.2) 206

§VII.6: Finality, Rationality, and the Unity of the Self 209

§VII.7: Kantian Unity 220

§VII.8: Korsgaard, Unity and the Bridge Function 222

§VII.9: Is the OP Necessary? 223

§VIII.10: Conclusion 229

Chapter VIII: The Great Unraveling 234

§VIII.1: The Content of Ideals 237

§VIII.2: Defending C_3 241

§VIII.3: Pluralism and the Failure of Congruence 248

§VIII.4: The Failure of Kantian Congruence 254

§VIII.5: The Great Unraveling 259

§VIII.6: Brief Contrasts with Other Accounts 266

Chapter IX: The Political Ideals of Justice as Fairness 270§IX.1: *PL*'s Basic Argument for Stability 273§IX.2: C_3' and the Sense of Justice 283§IX.3: C_3' and the Ideals of Conduct 287§IX.4: C_3' and the Social Ideals of Justice as Fairness 293

§IX.5: Whither Congruence? 296

Chapter X: Comprehensive Reasons to Be Just 301

§X.1: Moving from (9.2) and (9.3) to (9.5) 303

§X.2: Would there Be an Overlapping Consensus? 308

§X.3: Legitimacy and Justification 312

§X.4:	Why Political Legitimacy?	319
§X.5:	A Question about the Arguments for C_9 and C_{PL}	321
§X.6:	Public Reason, Mutual Assurance, and Pluralism about Justice	327
§X.7:	Stability, Reflective Equilibrium, and Public Justification	335
§X.8:	Conclusion	339
Chapter XI:	Conclusion: Why Political Liberalism?	344
§XI.1:	The Moral Basis of Political Liberalism?	347
§XI.2:	A Conception-Based View	353
§XI.3:	Defending Political Liberalism	357
§XI.4:	“And very good it was”	362
<i>Bibliography</i>		371
<i>Index</i>		375

Acknowledgments

This book has been a long time in the making and I have incurred many debts in the course of writing it. I first wrote up part of the book as a contribution to a conference organized at Trinity College, Dublin, in 2006 by Nigel Biggar. I received especially helpful comments at that conference from Nigel himself and from Nicholas Wolterstorff. For many years, both before and after the conference, I have conversed and corresponded with a number of people about Rawls's work. I have benefited greatly from my exchanges with Robert Adams, Robert Audi, Daniel Brudney, Peter de Marneffe, Neil Delaney, Thomas Pogge, Henry Richardson, Fred Rush, James Sterba, and Peter Wicks. In the spring of 2009, an audience at the University of Toronto heard a paper that distilled the central argument of the book. I am grateful to many of those who attended my talk, particularly Ronald Beiner and Melissa Williams, for helpful comments on that occasion. Alasdair MacIntyre read the manuscript of this book in its entirety and offered comments that were exceptionally trenchant and insightful. I owe a special debt to John Roos of the Notre Dame Political Science Department, with whom I have team-taught *A Theory of Justice* for many years, and to the many students who have passed through our course in Notre Dame's Philosophy, Politics and Economics Program. My understanding of Rawls's work would be much less precise if I had not had the privilege of working with John to explain it to those first-rate undergraduates.

It has been an honor and a pleasure to work with Oxford University Press. Comments by Colin Bird and a second, anonymous, reader did much to improve the book; indeed, one of Colin's comments led me to recast large parts of the argument in a form that I believe is considerably more perspicuous. Peter Ohlin of Oxford has been a patient and understanding editor. I am

very fortunate that Samuel Freeman inaugurated his series with Oxford just before I approached the Press about publishing this book and that Peter directed my manuscript to him. Samuel's comments at the final stages did much to improve the book. The book could not have reached publishable form without the patient labors of Natalie Johnson of Oxford and of Alex Jech. I am profoundly thankful for their help.

The long quotation at the beginning of Chapter VII is reprinted by permission of the publisher from *A Theory of Justice* by John Rawls, p. 574-75/503, Cambridge, MA: The Belknap Press of Harvard University Press, Copyright © 1971, 1999 by the President and Fellows of Harvard College.

Much of this book was drafted during an administrative leave I enjoyed after six years as chair of the Notre Dame Philosophy Department. My years of administrative work were, if not all-consuming, at least voracious consumers of my professional life. I am grateful to my colleagues for making those years such rewarding ones, and to the University for funding the leave that followed.

Sustained concentration on a book can impose an enormous strain on one's family. I am grateful to my late mother Patricia Weithman, to my beloved wife Maura Ryan, and to our wonderful children Annie and Meggie for bearing up under the strain without complaint, for overlooking the many things that I left undone while I was writing it, but most of all for their unfailing love.

When my colleague David Solomon first heard about this project, he described it as an exercise in filial piety. He meant that the book is a filial tribute by a Rawls student to his *doktorvater*, and in that he was right. But in describing the book as he did, David spoke more truly than he knew. My interest in philosophy was kindled and sustained by a succession of extraordinary teachers. Rawls and Judith Shklar, who codirected my dissertation, were among them and my debts to the two of them are incalculable, but many others have helped along the way. They include Karl Ameriks, Roderick Firth, Richard Foley, Warren Goldfarb, John Jenkins (now Fr. John Jenkins, CSC), Vaughn McKim, Ernan McMullin, and Martha Nussbaum. Mike Loux was my first teacher of philosophy, and no one could ask for a better start in the subject—though given Mike's ability to galvanize students, "jump start" would be a more apt description. David Solomon has an incomparably encompassing view of big pictures in the history of philosophy, and he has tried—with less success than he might hope—to get me to see and communicate them. Alvin Plantinga and Tim Scanlon taught me that depth of understanding is best won through uncompromising rigor.

Some of my former teachers are now my colleagues; most are, or were, my friends. It is a privilege to dedicate this book to my teachers with the deepest gratitude for all that they have done for me.

List of Tables

II.1	p. 48
II.2	p. 49
II.3	pp. 55, 92, 152
IX.1	p. 279

This page intentionally left blank

Why Political Liberalism?

This page intentionally left blank

Introduction

In the 1980s, John Rawls—author of the magisterial work *A Theory of Justice*¹—took what is sometimes described as a “political turn.” Justice as fairness, the conception of justice presented in *TJ*, was re-presented as what Rawls called a “political liberalism.” This re-presentation drew on a family of ideas and arguments that were new to justice as fairness, and reached its fullest expression in Rawls’s second major work, *Political Liberalism*.² In this book, I take up the important but underexplored question of why Rawls made the turn to political liberalism. Answering this question has a number of textual and philosophical payoffs. One is that it leads us to a fuller appreciation of the deep problems that Rawls tried to address by developing a theory of justice.

An explanation of Rawls’s turn to political liberalism should account for the differences between *TJ* and *PL*. Those differences are numerous and striking. I cannot discuss them all, and so it may help if I begin by listing those that I think stand in greatest need of explanation.

- In *PL*, the stability of a well-ordered society (WOS) is secured by an overlapping consensus of reasonable comprehensive doctrines.

1. John Rawls, *A Theory of Justice* (Harvard University Press, 1971 and 1999). I shall hereafter refer to this work as ‘*TJ*’ and cite it parenthetically in the body of the text. The first page references are to the 1971 edition, and the second are to the revised edition of 1999.

2. John Rawls, *Political Liberalism* (Columbia University Press, 1996). I shall hereafter refer to this work as ‘*PL*’ and cite it parenthetically in the body of the text.

- Justice as fairness is presented in that book as a political conception of justice, founded on basic ideas drawn from democratic political culture.
- In *PL*, the conception of the person represented by the original position—hereafter “the OP”—is said to be a political conception.
- The idea of public reasoning, which was hardly mentioned in *TJ*, is prominent in *PL*.
- The notion of political legitimacy, which received no explicit treatment in *TJ*, assumes a very prominent role in *PL*.
- In *PL*, Rawls admits that consensus in a WOS would probably focus on a family of liberal political conceptions of justice rather than on justice as fairness alone.
- *TJ*’s attempt to show that justice as fairness would be inherently stable is replaced in *PL* by an attempt to show that it would be stable “for the right reasons.”

Three other changes are less obvious but very important: Rawls’s description of the sense of justice and his argument that political society is a good undergo subtle but revealing changes, and the notion of congruence—so central to Rawls’s treatment of stability in *TJ*—does very little work in *PL*.

These are the changes in Rawls’s presentation of justice as fairness that I shall try to explain. Rawls made the changes to address shortcomings in the original presentation of his work. I take the position that Rawls thought the shortcomings he found were not merely shortcomings of interpretation, on his readers’ part, but were shortcomings in justice as fairness itself and—in particular—in its treatment of the stability of a WOS.

I have tried to offer periodic summaries throughout the book, and have provided numerous of cross-references. I therefore hope that the book will prove easy enough to navigate that I need not supply a detailed map or summary at the outset. Instead, I shall confine myself to a few remarks that will, I hope, provide a useful overview of the journey to come. The best way to furnish that overview may be to communicate the surprise that readers of this book, or parts of it, have expressed about the picture of Rawls’s work that emerges from it.

§1: Overview

A number of readers have said that the book introduces them to a very different Rawls than the one they thought they knew. Some of these readers still think of Rawls as a social choice theorist or a decision theorist. This book, they think, is not about the contractualist who once wrote that “the theory of justice is a part, perhaps the most significant part, of the theory of rational choice” (*TJ*, p. 16/15). Others have found this book surprising because they started with a quite different picture of Rawls. They think my claim that Rawls devoted considerable attention to avoiding collective action problems implies that he

was not the Kantian with whom they have become familiar. Collective action problems are indebted to one view of human rationality, they think, while Kantianism is animated by quite another. Still others have thought my argument that the justice of a well-ordered society depends upon large-scale changes in citizens' rational preferences shows that Rawls must be committed to a very non-Kantian account of moral motivation. Some readers have been surprised to meet a Rawls who has a persistent interest in the self and its unity. More have been surprised to meet a Rawls moved by deep questions about the goodness of humanity and the world.

These readers all started with something of the truth about justice as fairness. But as one reviewer of *TJ* said, "Rawls's theory has both the simplicity and the complexity of a Gothic cathedral."³ These readers' surprise shows that they missed a great deal by adopting just "one view of [that] cathedral"⁴ and by seeing Rawls's work from just one point of view. In this book, I try to develop and defend an interpretation that unifies their various perspectives and shows what truth there is in the various partial readings interpreters have extracted from Rawls's texts. I hope that the interpretation I defend is not only compelling, but also elegant and powerful in roughly the way that physical theories, economic theories, and mathematical results can be. Theories and results are elegant and powerful if they unify a lot on the basis of a little. I hope to do just that, showing how much of Rawls's work—including the most notable changes between *TJ* and *PL*—can be explained by supposing that he maintained a disciplined focus on a few intellectual concerns, and by seeing where those concerns led him.

One of Rawls's most pressing concerns was with the stability of a just society. He took up problems of stability in the third part of *Theory of Justice* and later in *Political Liberalism*. Seeing how Rawls initially thought he had shown that justice as fairness would be stable, and why he came to think that his original arguments for stability failed, shows why Rawls recast his view as a "political liberalism". By asking what Rawls means by 'stability' and what threats to stability he wanted to avert, we can unify the various perspectives on Rawls's work that I referred to a moment ago.

On my reading, Rawls wanted to identify basic terms of social cooperation that would be fair and collectively rational. Having identified those terms, he wanted to show that an arrangement which satisfied them would not be destabilized by a generalized prisoner's dilemma. At the same time, he wanted to show that they could be stabilized without reliance on a Hobbesian sovereign or a dominant ideology. Rather, he wanted the terms of cooperation to be

3. John Chapman, "Rawls's Theory of Justice," *American Political Science Review* 69, 2 (1975): 588-93, p. 588.

4. The phrase alludes to the title of Guido Calabresi and A. Douglas Melamed, "Property Rules, Liability Rules and Inalienability: One View of the Cathedral," *Harvard Law Review* 85, 6 (1972): pp. 1089-1128.

stabilized over time by the free activity of those who lived under them, in some robust sense of ‘free’

Rawls argued in *TJ*, and continued to believe in *PL*, that justice as fairness would be stable only if citizens in a WOS developed a sense of justice. He argued that they would. He also thought that justice as fairness would remain stable only if citizens of a WOS maintained their sense of justice. Maintaining a sense of justice requires a commitment to leading a certain kind of life. *TJ*’s treatment of what Rawls called “congruence” was supposed to show that members of a WOS would affirm and maintain their commitment to living justly, so that their sense of justice would be a standing element of their character.

In his *Lectures on the History of Moral Philosophy*, Rawls says Kant believed that an enduring good will may require “a kind of conversion” that is “strengthened by the cultivation of the virtues and of the ways of thought and feeling that support them.”⁵ The religious overtones of the word ‘conversion’ open the possibility that Kant thought the maintenance of a good will is a response to supernatural intervention in one’s life, a response that may need to be sustained by divine aid.⁶ Despite his affinities with Kant, Rawls clearly wanted to furnish a naturalistic account of how members of a WOS sustain their good will, or that ingredient of a good will that stability requires: their sense of justice. His argument that members of a well-ordered society would maintain their sense of justice therefore relies on a naturalistic psychology and, in particular, on a tendency to reciprocity that was, he conjectures, naturally selected for.

Because of this important feature of human psychology, Rawls argued that the “ways of thought and feeling” that support a sense of justice can be fostered by just institutions. Such institutions would shape the characters of those who live under them, so that they would respond in kind to benefits received, and would attach little value to what they could gain from free-riding and other forms of injustice. Caring little about these gains, they would not be drawn to plans of life that would leave them free to decide case-by-case whether to honor the principles of justice. Instead, they would adopt plans that would give their desire to honor the principles a central place. Because each member of the WOS would adopt such a plan, and would know that everyone else would do so as well, justice as fairness would be stable. Because the character formation necessary for stability would be effected by institutions that satisfy the principles of justice, and because those principles are the centerpiece of justice as fairness, Rawls concluded that justice as fairness—when institutionalized and publicized—would stabilize itself.

5. John Rawls, *Lectures on the History of Moral Philosophy* (Harvard University Press, 2000), ed. Barbara Herman, p. 155.

6. Patrick Freieron, *Freedom and Anthropology in Kant’s Moral Philosophy* (Cambridge University Press, 2003), p. 191, notes 31, 32, and 35.

Thus the Rawls of *TJ* recognized that an agreement reached in the original position could be undermined by a generalized prisoner's dilemma. Thinking he had shown that citizens of a just society would become the kind of persons who discount the pay-offs of injustice, he believed he had found a way to avert that threat without relying on a Hobbesian sovereign to alter citizens' pay-off tables. Furthermore, Rawls argued, because of the conditions of the original position, the principles that would be chosen there are principles members of the WOS would give themselves. And so when they regulated their lives by the principles, they would live lives that would be free in an important sense of 'free': they would live *autonomous* lives. Indeed, Rawls thought that one of the reasons they would endorse life-plans regulated by the demands of justice is that they would all want to live autonomously. Thus, *TJ*'s Kantianism was an essential part of Rawls' solution to the generalized prisoner's dilemma and his treatment of stability.

The possibility that members of the WOS would defect from fair terms of cooperation manifests a deep and familiar fact about human beings: we are creatures of divided hearts and wills. We can know what we should do and we can want to do it, but we can also be powerfully drawn to do something else—to advance our own interests, or those of people and causes we care about, in ways that are contrary to justice. This divide is a divide within our practical reason, a divide between what Rawls would come to call the Reasonable and the Rational. The stability of justice as fairness requires that our practical reason be unified and that our commitment to justice be—as Rawls would put it in *PL*—"wholehearted" (*PL*, p. xl). Because we are essentially reasoning beings, it requires that our selves be unified.

Few readers have recognized that *TJ*'s arguments for stability were intended to address the threat of a generalized prisoner's dilemma and to do so by showing how treating the principles of justice as regulative unifies human practical reason. If those arguments had succeeded, their success would have constituted a stunning philosophical achievement. Unfortunately, they did not. In the years following the publication of *TJ*, Rawls continued to accept his own earlier arguments that members of a WOS would develop a sense of justice, though in *PL* he made some important changes that he failed fully to acknowledge. But he came to realize that his argument that members of the WOS would maintain their sense of justice failed, and with it, his argument that a WOS would not be destabilized by a generalized prisoner's dilemma. And so he came to realize that he needed to offer a different set of arguments for those conclusions. Offering those new arguments required Rawls to recast justice as fairness as a political liberalism. The changes between *TJ* and *PL* that I listed above can be explained by seeing how they facilitate those new arguments.

Rawls's arguments for stability, both early and late, depend upon our natural amenability to developing a sense of justice and our natural amenability to the other developments of our character that just institutions are supposed to bring about. We can be naturally amenable to these developments

only if we have what the Rawls of *PL* called a “moral nature.” By that he meant “not . . . a perfect such nature, yet one that can understand, act on and be sufficiently moved by a reasonable political conception of right and justice[.]” (*PL*, lxii) And so I believe Rawls thought that we can be amenable to the requisite moral development only if we are, or under the right circumstances can become, good. The arguments for stability in *PL*, if sound, vindicate the claim that we can be. If we are at least capable of being good, then—however we may actually behave—our presence in the world need not mar creation. The upshot, as I shall argue in the Conclusion, is that Rawls’s theory of justice can be read as a brilliant and subtle exercise in naturalistic theodicy. Rawls offers arguments one consequence of which is that, despite the evil for which human beings are responsible, a good Creator could still have seen fit to fashion a world with us in it.

§2: The Road to Come

I have sketched my interpretation in broad strokes to provide readers some orientation, but the journey that follows goes by way of considerable textual and philosophical detail. According to the reading put forth here, Rawls took his political turn because there were clearly identifiable arguments in the original presentation of justice as fairness with which he later became dissatisfied. We can explain the changes between *TJ* and *PL* only by locating those arguments, laying them out with care, supplying missing premises when necessary, and asking where Rawls might have thought those arguments went wrong. We can then pinpoint key premises he came to reject as implausible, and others that he modified to facilitate his political turn.

I am not, of course, the only reader of Rawls who thinks we need to look at shortcomings of argument to find reasons for his political turn, but my reading of Rawls’s reasons for the turn to political liberalism stands in sharp contrast to the interpretation that I think is most popular. That interpretation, which I call the *Public Basis View*, locates the shortcomings in an argument for the principles of justice that is said to be implicit in part I of *TJ*. That argument for the principles, which I call “the Pivotal Argument,” is itself of considerable interest and serves as a useful analytic device to which I shall return periodically throughout the book. I therefore take some pains to lay it out precisely in Chapter I. Once the argument is laid out, the *Public Basis View* can be seen to have considerable appeal. I shall argue, however, that it founders on textual and philosophical shortcomings that prove insuperable.

I have said that the arguments with which Rawls became dissatisfied are to be found in the part of *TJ* devoted to the stability of justice as fairness and, in particular, in *TJ*’s treatment of congruence. In Chapter II, I distinguish various kinds of stability and identify the kind in which the Rawls of *TJ* was most interested—what he referred to as “inherent stability.” Chapter II also

identifies, more clearly than is often done, the threat to stability with which Rawls was concerned. As I have already indicated, showing that justice as fairness would be inherently stable required showing that it could, when institutionalized, survive the threat of the generalized prisoner's dilemma without relying on a Hobbesian sovereign.

Chapters II and III show, in general terms, that *TJ*'s argument for the congruence of justice and goodness is a crucial part of Rawls's larger argument that justice as fairness would survive that threat, and so would be inherently stable. The problem with *TJ*'s treatment of stability, Rawls came to think, was that it relied on the improbable assumption that members of the WOS share what he called a "comprehensive doctrine." In Chapter III, I spell out what Rawls means by "a comprehensive doctrine," what he means by "congruence," where he thought his treatment of congruence relied on the assumption about a shared comprehensive doctrine that he later found implausible, and why reliance on that assumption in *TJ* led to an inconsistency in justice as fairness.

Some of the best published literature that treats of Rawls's congruence arguments mistake the structure of the congruence of arguments, the sequence of arguments that are offered, and the ways in which the various congruence arguments hang together. I give a good deal of attention to reconstructing those arguments, since I think we will see where Rawls thought the arguments went wrong only if we first see how he originally intended them to go. Chapter IV lays the groundwork for those arguments by attending to the acquisition of the desires they presuppose. Chapters V through VII lay out the arguments. In Chapter VIII, I go through the steps by which Rawls's treatment of congruence—so carefully knitted together in *TJ* and, as we shall see, in the original *Dewey Lectures*—came unraveled.

In Chapters IX and X, I show how the changes introduced between *TJ* and *PL* respond to the difficulties Rawls found in *TJ*'s treatment of stability. In the conclusion, I answer the question that gives this book its title by defending political liberalism against a common but powerful objection, by contrasting justice as fairness with another version of political liberalism, and by showing how political liberalism helps to answer the questions about the goodness of humanity and the world that, I have said, concerned Rawls so deeply.

§3: A Deeper Understanding of Justice as Fairness?

The congruence arguments in *TJ* are laid out in a single section late in the book. The claim that Rawls took his political turn because of problems in his original treatment of congruence might be thought to suggest the implausible thesis that Rawls made very far-reaching changes in his view because of shortcomings in a couple of pages of argument. In fact, as we shall see, the problems that Rawls identified in his treatment of congruence go to the heart of his

constructivism. That is one of the reasons Rawls came to think that the repairs needed by justice as fairness had to be so extensive. Moreover, the congruence arguments, when properly reconstructed, are seen to draw on material and concerns from throughout *TJ*. Making explicit how they did so brings some of the concerns and structure of *TJ* to light. One thing that is apparent from the recovery of the congruence arguments, for example, is that Rawls's concern with intuitionism—which he seemed to dispatch by the end of *TJ*, §7—was much more profound and pervasive than it is usually thought to be. Another is that the ambitious but puzzling discussion of the unity of the self in *TJ*, §85 responds to Rawls's deep and abiding concerns about how practical reason is to be unified. Appreciating that section, I believe, deepens our appreciation of the Kantian Interpretation of justice as fairness laid out in *TJ*, §40. It especially heightens our appreciation for the crucial role Rawls assigned a Kantian conception of the person in *TJ*.

As these remarks suggest, one striking feature of the treatment of congruence is the extent to which it draws on other sections of *TJ*, and on other sections of part III in particular. One of the reasons we learn so much about *TJ*, and about justice as fairness, by asking why Rawls turned to political liberalism is that we come to see how parts of *TJ* fit together, in unanticipated ways, by making the congruence arguments explicit. Part III of *TJ* is sometimes read as if it were an undisciplined attempt to cover some of Rawls's favorite topics in ethics. The material on the moral and natural sentiments, for example, can appear to be set of tangential arguments directed against crude forms of emotivism and prescriptivism. In fact, I believe part III is exemplary for the way it painstakingly establishes conclusions with an eye toward their later use in Rawls's arguments for stability. We shall see that the continuity of the sentiments is crucial for the second congruence argument Rawls offers in *TJ*, §86. While this book is not a commentary on part III of *TJ*, I hope it will go some way toward rekindling interest in that neglected part of the book.

Pursuing the reasons for Rawls's political turn also puts us in a position to see how much of the treatment of moral development in *TJ*, chapter 8 survives the transition to *PL*. This is a natural question to raise about justice as fairness, since Rawls rarely spoke of a sense of justice after *TJ* and did not return to the process of moral development in any systematic way. But I do not think that that is because other matters eclipsed his concern with the development of a sense of justice or because he thought his discussion of moral development needed to be abandoned. Rather, as I hinted earlier, Rawls continued to think the question of whether a WOS would be stable had a two-part answer. The first part was provided by showing that members of the WOS would develop a sense of justice. The second was provided by showing that they would judge that preserving their sense of justice belongs to their good. Rawls did not revisit *TJ*'s treatment of the first part in subsequent work because, he says, he continued to think it was adequate, and could survive the changes in his view. Rawls made the changes between *TJ* and *PL* because he thought they were necessary to support the second part of the answer; I shall suggest that, his

claims to the contrary notwithstanding, Rawls himself thought changes in the first part—at least changes of emphasis—were called for as well.

An especially important question about justice as fairness concerns the dispensability of the original position. That question has hung over Rawls's work for almost four decades. Rawls's insistence in his later work that the original position is a device of representation seems to invite the question in urgent form, but that question was pressed in some quarters well before the political turn. I argue that the original position is a theoretical device that "bridges" the right and the good in Rawls's early work, for it functions in the argument by which Rawls identifies principles of right and in an argument by which Rawls argues that acting from those principles belongs to the good of members of the WOS. The original position may not be necessary for the first argument but, I shall argue, it is necessary for the second. The second argument was, in turn, necessary to solve the question of congruence in *TJ* and the *Dewey Lectures*. The original position is not, therefore, dispensable from the arguments Rawls offered for justice as fairness before his political turn.

Perhaps the most notable feature of Rawls's re-presentation of justice as fairness is its starting point. Rawls insists that as a political liberalism, justice as fairness begins with ideas and convictions latent in the public political culture of liberal democracy. Most readers have considered this to be a marked—if not a revolutionary—change from the philosophical method of *TJ*. Some, as we shall see, have accused Rawls of moral retrenchment. I shall argue, against the conventional wisdom, that even in *TJ*, Rawls took for granted a view that members of liberal democratic societies can normally be expected to have of themselves, and that in the course of developing justice as fairness he refined that view of the person and gave it a central role.

Thus even before his political turn, Rawls started from within—and addressed his work to—the liberal democratic world. The difference between his earlier and later presentations of justice as fairness is not, therefore, that the latter starts within that world while the former does not. The difference lies in what he drew from liberal democratic culture. In his early work, it was an ethical—not a metaphysical—conception of the person, a conception that he further specified in ways that he came to think could be an object of controversy among reasonable citizens. In his later work, he was made clear that the conception of the person he drew from political culture was a specifically political conception.

§4: Unity, Theodicy, and the Attractions of Liberalism

By looking closely into why Rawls made the changes between *TJ* and *PL*, we also learn a great deal about liberalism, its attractions, and its ambitions.

The theoretical foundation of liberalism is sometimes said to be a set of rights or a basic right, such as the right to equal concern and respect. That is

why some readers, most famously Ronald Dworkin, interpret Rawls's liberalism as rights-based. Though Charles Larmore has argued that a principle of legitimacy lies at the core of political liberalism, he thinks that what the principle of legitimacy really expresses is an imperative of respect for persons, and so his reading has strong affinities with Dworkin's.⁷

The role of reflective equilibrium in justifying justice as fairness implies that there is some artificiality to speaking of a "foundation" for Rawls's liberalism. Those qualifications notwithstanding, the reading of Rawls that I defend here shows that justice as fairness is an alternative to rights-based—and hence to legitimacy-based—theories of justice. On my reading, Rawls supposes from the outset that under the impact of liberal democratic thought and practice, we, his readers, think of ourselves as free and equal persons embedded in a society that ought to be a fair scheme of social cooperation. We have, he thinks, a democratic conception of our society and a conception of ourselves that I call a *free-and-equal self-conception*.

Crudely put, Rawls refines and specifies these conceptions so that they yield an answer to the question he poses in the *Dewey Lectures*: what conception of justice is best suited to regulate the collective political life of persons who think of themselves as free and equal members of a fair cooperative scheme? Liberal rights, and a liberal conception of legitimacy, are not the foundations of his liberalism, though they are part of Rawls's answer to that question. As we shall see, his principle of legitimacy, as stated in *PL*, is justified by showing that our exercises of political power must conform to that principle if we are to live as free and equal persons, properly conceived, and to enjoy what I shall call the *Ideal of Democratic Governance*. Thus, if we can speak of the "foundation" or "foundations" of justice as fairness at all, what is foundational to it are conceptions of the person and of society that are found in democratic culture and that are made specific enough to generate political principles. Justice as fairness therefore illustrates—as Rawls himself says—the possibility of a liberalism that is "conception-based" or "ideal-based," rather than "rights-based."⁸

The attraction of Rawls's principles of justice depends in part upon their distributive implications. But it also depends on the attractiveness of the political conception or ideal of the person on which they are based, for among the reasons we have for acting from the principles is that by doing so, we will realize that ideal. That ideal is, I believe, very attractive. Its attractiveness is important. Some critics, put off by what they see as the individualism, selfishness, and materialism of modern life, claim that liberalism invariably

7. See Charles Larmore, "The Moral Basis of Political Liberalism," in Larmore, *The Autonomy of Morality* (Cambridge University Press, 2008), pp. 139–67, especially pp. 146ff.

8. John Rawls, "Justice as Fairness: Political not Metaphysical," in John Rawls, *Collected Papers*, (Harvard University Press, 1999), ed. Samuel Freeman, pp. 388–414, pp. 400–401, note 19; Rawls credits Elizabeth Anderson with describing his view as "ideal-based."

produces the kind of person they deplore. They defend other forms of political life as better suited to our social nature. One way to answer these critics is to show that liberalism does take due account of our social nature, and encourages us to live up to conceptions of ourselves that lack the features on which critics seize.

Rawls's liberalism suggests how this might be done. Rawls is often read as propounding an individualistic theory. The argument for the principles, which relies on the device of a social contract, can be described that way. But according to *TJ*'s arguments for congruence, members of the WOS would judge that upholding the principles is part of their good because it is only by upholding the principles that they can satisfy natural desires for friendship, association, and sincere and open dealings with others. Though Rawls modified those arguments considerably in his later work, he continued to think that part of what makes his principles attractive is that acting from them enables us to live among others in ways that should appeal and inspire.

Some readers have said that on reading *TJ*, they thought that their own deepest moral convictions had received their best expression and their most powerful defense. Others of us had a somewhat different reaction. Justice as fairness expressed our deepest political convictions. But we came to political philosophy with deeply held views about what is good in life and why, and those conceptions of the good had implications for the right that were not obviously compatible with justice as fairness. The result was a tension between potentially conflicting identities.

In Rawlsian terms this tension reflects a conflict between the demands of conceptions of justice associated with our views of the good, on the one hand, and the demands of the Reasonable on the other. The attraction of justice as fairness is not, therefore, the attraction of something that is alien to those who have traditional views of the good. It is the attraction we feel for the reasonable part of ourselves. Rawls's concern with the unity of the self showed the tremendous ambition of *TJ* and promised to show how the tension should be resolved. For Rawls argued that the only way creatures like us can live as unified selves, at least under modern conditions, is to regulate our pursuit of the good by principles of liberal democratic justice. The alternative to being regulated by the reasonable part of ourselves was, Rawls seemed to suggest, to live lives that lacked rational unity. That is why—though Rawls had said of the parties in the OP that their aim “is to establish just and favorable conditions for each to fashion his own unity” (*TJ*, p. 563/493)—he also said that what he called the “essential unity” of the self is established by taking the principles of justice as supremely regulative (*TJ*, p. 563/493).

An important part of the congruence argument, I will suggest, is devoted to establishing this last claim. We shall see that one of the reasons Rawls became dissatisfied with his treatment of congruence was that he realized a truly liberal view cannot take a stand on how the “essential unity” of selves is to be attained. And so while he continued to think that each citizen in the WOS would treat the principles of justice as in some sense regulative, he also

came to recognize that how the principles of justice are to be connected with or founded on various conceptions of the good must be left to each person to work out. I believe that one reason for taking Rawls's principles as regulative of our political lives is the great attraction of being the kind of citizens justice as fairness calls us to be. Seeing that we can be that kind of citizen, in turn, completes what I referred to earlier as Rawls's "naturalistic theodicy," for it vindicates our hope in the possibility of a world that is more just and that can aptly be called "good."

§5: A Final Word to the Reader

I have given some indication of what I think can be learned by pursuing questions about why Rawls made the changes he did between *TJ* and *PL*. I conclude this introduction by saying a few words about what I shall ask of readers and about the limitations of the book.

As my remarks so far have suggested, this book is not intended as a primer in the main lines of Rawls's thought. Moreover, at this point, the literature on Rawls is so well developed, and the study of his work so widespread and thorough, that I feel justified in presupposing an acquaintance with the major ideas and texts that is fairly sophisticated. A sign of the familiarity that I presuppose is that I use abbreviations like WOS for "well-ordered society" and OP for "original position." Because Rawls's texts and ideas have attracted so much critical attention, I also assume that any reading that hopes to offer something new must be very carefully defended and very firmly anchored in the text. I have therefore hewn closely to the written word and used an expository style that is more commonly found in other areas of philosophy, spelling out some of Rawls's reasoning in premise-and-conclusion form. Some of Rawls's arguments compress very complicated lines of thought and, as I have already implied, the compression in *TJ* is facilitated by Rawls's frequent reliance in one argument on conclusions that have been established by other arguments elsewhere in the book. The reconstructions that I provide can therefore be demanding. I have made demands of readers because I believe the reconstructions heighten appreciation for the rigor of Rawls's own arguments, and that the method of exposition I have chosen makes analysis of those arguments more economical and perspicuous.

Some of the most demanding reconstructions are in Chapters IV through VII, where *TJ*'s congruence arguments are laid out and analyzed. Chapter VIII, which tells why Rawls became dissatisfied with those arguments, depends upon the chapters that immediately precede it. These four chapters together supply the interpretation offered here with some of its most detailed textual and philosophical support. As I have already indicated, Chapters II and III provide an overview of *TJ*'s treatment of stability and of the reasons Rawls became dissatisfied with it. Readers who are less interested in the details of the

congruence arguments, who are uninterested in textual exegesis, or who are content with a general understanding of why Rawls made the turn to political liberalism, are invited to read selectively between Chapter III and Chapters IX and X. There, I show how the changes Rawls made after *TJ* respond to the sources of his dissatisfaction with his earlier arguments.

This book is intended to be a defense of political liberalism, but it is a defense of an unusual kind. Though I do reply to some standard objections to political liberalism in the Conclusion, the book is not an attempt to defend Rawls's later views against all comers. Rather, the defense provided here is the kind of defense Gerald Cohen hoped to provide of Karl Marx's theory of history—a defense that proceeds “by offering argument in its favor, but more by presenting the theory in what I hope is an attractive form.”⁹ While I did not face the challenge that Cohen did, I thought that one attractive form in which political liberalism still needed to be presented is as a rigorous and systematic response to a specific set of problems which Rawls correctly came to see in premises and arguments on which he had previously relied. I hope that my end is served by the care with which I have tried to lay out Rawls's lines of thought, both early and late, and by my attempt to display the underlying unity of his views.

I am strongly inclined to think that Rawls succeeded at what he set out to do: identify fair and collectively rational principles of justice that, when institutionalized and publicized, avert the threats to stability with which I have claimed he was concerned. Unfortunately, laying out and unifying Rawls's treatment of stability within tolerable bounds of length meant giving less critical scrutiny to certain crucial claims than I would have liked. There are many places at which what Rawls says admits of more than one interpretation, at least when what he says is taken in isolation. Quite often, I have assumed that readers of Rawls will already have noticed the ambiguity and that my job is to stake out a position on an interpretive question rather than to belabor the way the question arises. In these cases, I have opted for what I take to be the best reading and shown that it makes sense of the larger argument, without explicitly distinguishing and puzzling through the various interpretations the text will bear.

As I have indicated, Rawls's arguments for stability depend upon psychological assumptions. Those assumptions need probing. One assumption, or set of assumptions, is especially in need of attention: Rawls's assumption that acquisitiveness has its origins in the desire for status. This assumption does considerable philosophical and political work in justice as fairness. It is an assumption Rawls held throughout his working life.¹⁰ In §V.4, I have tried

9. Gerald Cohen, *Karl Marx's Theory of History: A Defence* (Princeton University Press, 1978), p. ix.

10. See my review of John Rawls, *A Brief Inquiry into the Meaning of Sin and Faith* (Harvard University Press, 2009), ed. Nagel, *Notre Dame Philosophical Reviews*, <http://ndpr.nd.edu/review.cfm?id=17045>

to understand why the assumption might hold in a special case and I have expressed some skepticism about it elsewhere,¹¹ but I have not subjected it to anything like the attention it deserves. Unfortunately, that will have to await another occasion. My aim has been to convey a synoptic view of how and why Rawls rebuilt his cathedral; doing so left me less scope than I would have liked to test this particular buttress.

Academic work is a way of serving others. I recognize that this book may be of greatest service to those who have wrestled with Rawls's texts for a long time, who remain puzzled about how certain of his arguments go and who wonder what he could have meant by certain obviously crucial but vexing assertions and turns of phrase. Even after some decades of scholarly attention to Rawls's work, I believe there is still a need for a book that pays attends so closely to texts that bear on his political turn and that tries to figure out exactly how his arguments go. That is the need I have tried to fill here. But I hope that this book will also be of service to all those who wonder whether a just world is possible, whether we human beings are capable of sustaining such a world, and whether those of us with traditional conceptions of the good can achieve some unity of self while living with others as free equals under modern conditions. These questions were, I believe, of the deepest concern to the greatest political philosopher of our time. In writing this book, I have tried to understand how he posed and answered them.

11. In my review of Rawls, *Brief Inquiry*; also in my "John Rawls and the Task of Political Philosophy," *The Review of Politics* 71 (2009): pp. 113–25.



The *Public Basis View*

Rawls made the changes between *TJ* and *PL* because he became dissatisfied with arguments that were critical to the presentation of justice as fairness in his first book. Any serious attempt to explain those changes must therefore identify the arguments with which Rawls became dissatisfied and say why he came to think they were unsatisfactory. In Chapter II, I shall say what I think those arguments were and what problems Rawls found with them. My account of the changes between *TJ* and *PL* challenges what I take to be the standard explanation of those changes. I shall refer to that explanation as the *Public Basis View* of the changes, and I shall devote this chapter to laying it out and evaluating it.

The label I have attached to the *Public Basis View* is new, but I think the *View* itself is widely accepted. Indeed, I believe that most readers who have an opinion about why Rawls introduced the changes between *TJ* and *PL* accept the *Public Basis View* in some form. I shall begin by developing the *Public Basis View* of the changes as an ideal type. I believe that the essentials of the *View* will be recognizable to those familiar with literature about, and discussion of, Rawls's turn to political liberalism. Later, I shall suggest that some philosophers who have developed prominent political liberalisms of their own endorse the *Public Basis View* of Rawls's political turn.

§I.1: Initial Statement of the *Public Basis View*

The *Public Basis View* of Rawls's transition is most easily explained and made vivid by relying on a certain picture of Rawls's WOS—a picture according to

which the WOS has a public charter that is expressed in fundamental political documents which play roughly the role in that society that the Declaration of Independence and the Constitution play in American political culture. While Rawls himself may not have had that picture in mind, it is not out of the question that he did and, as I hope will be evident, the picture has some heuristic value.

Rawls says that in the WOS of *TJ*, everyone would accept and would know that everyone else accepts the same conception of justice—just as, in the United States, citizens recognize and know that others recognize the rights and liberties accorded everyone by the Constitution. It is that conception, Rawls says, that serves as the WOS’s “foundation charter” (*TJ*, p. 11/10). By that Rawls meant that it was to serve as the shared, public basis for distributing benefits and burdens of social cooperation. If justice as fairness were to serve as a *shared* basis of justification, then it would have to be defended with an argument or a set of arguments that could be affirmed by all members of the WOS, so that everyone would accept the same principles of justice and accept them on the same grounds. This is the sort of defense Rawls hoped to provide in part I of *TJ*.

In the WOS of justice as fairness, the defense of the principles would be publicly available in important documents, just as the philosophical justification of American government is alluded to in the Declaration of Independence. That justification is alluded to in the second paragraph of the Declaration, which famously begins:

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness.—That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed[.]

Thus, according to the publicly available foundation of the American government, the ends, limits, and powers of government are justified—via consent—by a conception of the person. Similarly, according to the *Public Basis View*, the publicly available justification of justice as fairness would justify *it*—via consent—by a metaphysical conception of the person. In the Founders’ United States, the publicly articulated, metaphysical conception of human beings asserts that we are created free and equal by God. So in the WOS of *TJ*, it might be thought, the publicly articulated metaphysical view of human beings would be or would seem to be the Kantian view of human autonomy and equality expressed in the original position.¹

According to the *Public Basis View*, the problem Rawls came to see grows out of the fact that the WOS of justice as fairness would be a liberal society. Its

1. For Rawls’s exposition of the Kantian interpretation of justice as fairness, see Rawls, *TJ*, §40.

members would be free to explore and adopt a variety of religious and philosophical views about the good—what Rawls later labeled “comprehensive views” or “comprehensive philosophical doctrines” (*PL*, p. xviii). As Rawls came more deeply to appreciate this “fact of reasonable pluralism” (*PL*, p. 36), he came to see that the Kantian conception of the person is not a neutral starting point for political theorizing, but is a conception with which many reasonable people in a pluralistic society would disagree. At the same time, it is said, critics like Michael Sandel showed just how heavily the original position argument for the two principles depended upon the contentious metaphysical conception of the person from which Rawls began. Rawls therefore realized that reasonable people in a pluralistic society might reject the metaphysical argument he provided—or could be read as providing—for his principles. Since the WOS of justice as fairness would be a pluralistic society of reasonable people, he came to realize that the WOS might not be one in which everyone accepted the same conception of justice and its public defense after all.

To remedy this tension or seeming tension in his view, proponents of the *Public Basis View* claim, Rawls recast his defense of the principles so that it rested on premises that could be accepted by citizens who adhered to a wide variety of conceptions of the good and of the person—premises that were compatible with those conceptions because they were “political not metaphysical.” The public defense of justice as fairness was then explicitly said to begin, not from a metaphysical conception of the person, but from the conception of the citizen found in the public political culture of a democratic society. The principles of justice were then said to be justified—via consent—by this political conception of the person. The political premises of the new defense could then serve as the shared, public basis of the principles that Rawls had hoped to provide in *TJ*. Because members of the WOS endorse those premises from within their own comprehensive doctrines, the “foundation charter” of the WOS is, as it were, an area of “overlap” among otherwise divergent doctrines—hence the image of an “overlapping consensus.”

This brief summary of the *Public Basis View* may exaggerate—or may draw out at greater length than any proponent of the *View* would—the parallels between the Declaration of Independence and the public defense of justice as fairness in the WOS. But by doing so, it makes vivid three of the central claims of the *Public Basis View*: (i) the claim that the argument with which Rawls became dissatisfied was the argument for the principles of justice provided in part I of *TJ*, (ii) the claim that Rawls became dissatisfied with it because he recognized that it would be too controversial to serve as the shared, public basis of the principles in a pluralistic society, and (iii) the claim that Rawls responded to this difficulty by recasting that defense so that it could be the object of an overlapping consensus.

In one respect, however, the summary is too simple, since it suggests that there is a single *Public Basis View*. But at a critical juncture in the summary, I said that according to the *Public Basis View*, Rawls came to realize that he “provided – or could be read as providing” a defense of his principles that relied

upon a metaphysical conception of the person. This disjunction suggests two different reasons for Rawls's dissatisfaction with *TJ*'s defense of the principles of justice. There are therefore two different versions of the *Public Basis View*, which I shall refer to as the "strong" and "weak" versions.

Proponents of the strong version claim that Rawls's defense of the principles of justice really did rely upon metaphysical claims about persons. In moving from *TJ* to *PL*, they say, he disavowed those claims in favor of other arguments for the principles, arguments the premises of which are "political not metaphysical." Thus, in its strongest form, the *Public Basis View* is a thesis about substantive changes in justice as fairness, which involve the rejection of some metaphysical claims that Rawls previously endorsed. It is now widely thought that the central contention of this version is mistaken, for *TJ*'s argument for the principles of justice is now thought not to depend upon metaphysical claims. Even if this is so, there are two reasons why the strong version of the *Public Basis View* remains worthy of attention. One is that it is instructive to see just what is meant by denying that Rawls relies on metaphysical claims, since—though this is not generally appreciated—I think Rawls himself had something fairly precise in mind in denying it. The other is that the failure of the strong version of the *Public Basis View* suggests the weaker—and hence the more broadly appealing—version of the *View*.

Proponents of the weak variant recognize that many readers of *TJ*—including proponents of the strong variant—took Rawls's defense of the principles to depend upon metaphysical assumptions. But they deny that Rawls ever meant the premises of his defense to be taken this way. They think Rawls took an explicitly political turn in order to make clear that this metaphysical reading of those premises was wrong. The new ideas introduced in *PL*—such as the ideas of an overlapping consensus, the political conception of the person, and political autonomy—are said to be ideas Rawls introduced to explain what he meant all along.

At the heart of the both versions of the *Public Basis View* is, of course, the argument for the principles with which Rawls is alleged to have become dissatisfied—because it either relied on metaphysical claims or seemed to rely on them. I shall offer a concise version of that argument in the next section. Since that argument is, as it were, the pivot around which he is said to have made his political turn, I shall refer to that argument as the "Pivotal Argument." In order to see the appeal—and what I shall argue are the fatal textual and philosophical shortcomings—of the *Public Basis View*, it is necessary to go beyond the rough statement of the *View* I have given in this section and to lay out that argument rigorously. Some of the steps are unfortunately rather cumbersome, but having the argument before us will make for economy and clarity later on, since I shall refer to some of the steps frequently in the chapters to come. I shall not contend that *Public Basis View* is mistaken in supposing that Rawls relied on the Pivotal Argument or on an argument very like it, nor shall I deny that Rawls modified certain key claims in the argument as part of his transition to political liberalism. About these things, the *Public Basis View* is importantly

right. What I do deny is that the Pivotal Argument is the argument with which Rawls primarily became dissatisfied after publishing *TJ*. The changes Rawls made in his defense of the principles were motivated by his dissatisfaction with—and his need fundamentally to rethink—a very different set of arguments, found in a different part of *TJ*.

§1.2: The Pivotal Argument

I said earlier that I am developing the *Public Basis View* as an idealized position with which to contrast my own explanation of Rawls's political turn. The Pivotal Argument is not, therefore, an argument that is explicitly attributed to Rawls in any one article of scholarly literature. Rather, it is an argument that has to be supplied as part of the rational reconstruction of a view about changes between *TJ* and *PL* that is widely, if implicitly, held. In this section, I attempt to supply it.

When I sketched the *Public Basis View* in the last section, I implied that the Pivotal Argument follows a sequence of thought that begins with an assertion about human nature and proceeds, via consent in the original position, to Rawls's two principles. What I have called "the Pivotal Argument" therefore begins with a claim about human nature:

- (1.1) We are by nature free and equal rational agents who can reflect upon the ends we pursue, and can assess social arrangements in light of our own interests and ends.

Rawls assumes that human beings need access to the primary goods regardless of what ends they adopt. Those goods are produced and distributed by the basic structure of society. Because access to these goods is necessary—and because these goods are distributed by institutions whose influence is pervasive—our life prospects, our aspirations, and our sense of what is just and unjust, all are deeply affected by the distribution of primary goods. This gives us a powerful interest in how primary goods are distributed. And so the second step in the Pivotal Argument is:

- (1.2) We have a fundamental interest in the ways the basic structure of our society distributes the primary goods.

The fundamental interest we have in the production and distribution of primary goods makes their production and distribution a matter of justice. And so the principles in accord with which the basic structure produces and distributes primary goods must conform to what justice demands.

The task of determining what justice demands of the basic structure is, of course, the task Rawls sets himself in *TJ*. He locates his attempt to answer that question squarely in the contract tradition. Like others in the contract tradition, Rawls seems to make a crucial assumption about how basic social

arrangements are determined. Crudely put, forcing people to live under arrangements that are not acceptable to them is inconsistent with respecting them as the kinds of beings (1.1) says they are. More precisely:

- (1.3) If we have a fundamental interest in basic social arrangements, and if we are capable of rationally assessing those arrangements in light of our interests, then respect for us as free and equal persons with that interest and capability requires that the principles governing those arrangements be acceptable to us as such persons.

(1.3) is a conditional. The consequent is conditional on the claim that persons have any fundamental interest in basic social arrangements at all. One such interest they would have is an interest asserted in (1.2), the interest in how the basic structure produces and distributes primary goods. So (1.3) seems to imply that:

- (1.4) If (1.2) is true, and if we are capable of rationally assessing the ways the basic structure distributes primary goods in light of our interests, then respect for us as free and equal persons with that interest and capability requires that the principles governing the basic structure be acceptable to us as such persons.

I have already argued for (1.2). And (1.1) implies that we are capable of rationally assessing the way the basic structure produces and distributes primary goods. So (1.1), (1.2), and (1.4) imply:

- (1.5) Our society respects us as the kind of persons (1.1) says we are only if the principles governing the ways the basic structure of our society distributes primary goods are acceptable to us as such persons.

If Rawls also assumes that persons must be respected by their society as the kind of being (1.1) says they are then, since (1.1) says we are free and equal persons, the assumption that we must be respected—together with (1.5)—implies that:

- (1.6) The principles governing the ways the basic structure distributes primary goods must be acceptable to us as free and equal persons.

What does it mean to say that principles are or are not *acceptable* to us? And what does it mean to say that they are or are not acceptable to us *as free and equal persons*?

To say that principles are acceptable to us is to say that, if given the choice, we would accept them. To say that principles are acceptable to us *as free and equal persons* qualifies or elucidates the conditions under which they must be accepted. A crucial move in the Pivotal Argument is the claim that if the principles that govern distribution among persons were determined by features of their situation that are irrelevant from a moral point of view, then those persons would not really be treated as equals, since equal treatment requires leaving such considerations aside. This assumption requires that those who choose

or accept the principles must determine the principles free of the influence of those contingencies. And so:

- (1.7) The principles governing the ways the basic structure distributes primary goods must be acceptable in a choice situation that is uninfluenced by natural and social contingencies.

Once these contingencies are screened out, what is decisive in determining what principles we would accept is our nature as persons. There is nothing else left to determine the choice. So (1.7) implies:

- (1.8) The principles governing the ways the basic structure distributes primary goods must be acceptable in a choice situation in which our nature as free and equal persons is the decisive determining element of the choice.

The first premise of the Pivotal Argument, (1.1), is a claim about what we are by nature: free and equal rational agents capable of reflecting on ends and assessing social arrangements in light of our interests. This is just the way that we are represented in Rawls's choice situation, the OP. Indeed, it seems, the OP is constructed precisely so that nothing other than our nature as described in (1.1) affects what principles are adopted there. So Rawls seems to think that:

- (1.9) The OP is a choice situation in which our nature is the decisive determining element.

From (1.8) and (1.9), it follows that

- (1.10) The principles governing the ways the basic structure distributes primary goods must be acceptable in the OP.

Acceptability in the OP is determined by a series of pair-wise comparisons. And since Rawls argues that his two principles would be chosen in preference to other principles in the OP, he concludes that:

- C_1 : The distribution of primary goods by the basic structure must be governed by the two principles.

This is the Pivotal Argument. It is the line of thought by which the *Public Basis View* alleges that Rawls's principles would be publicly justified in the WOS of *TJ*. It is also the line of thought with which readers sympathetic to the *View* allege that Rawls became dissatisfied.

§1.3: Imputing the Pivotal Argument?

The plausibility of the *Public Basis View* depends upon the plausibility of imputing the Pivotal Argument to Rawls. There are some textual bases for imputing it.

Some of those bases were canvassed by Michael Sandel. Sandel famously went to some lengths to argue that Rawls defended his principles of justice by

relying on claims about persons that Sandel interprets as metaphysical.² Sandel was undoubtedly right that there is a conception of the person at work in *TJ* according to which members of the WOS are as (1.1) describes them. Sandel was also right to claim that the work done by that conception includes shaping the OP. For in the original edition of *TJ*, Rawls says that “the desire for liberty is the chief regulative interest that the parties must suppose they will all have in common in due course” and that the veil of ignorance “lead[s] to this conclusion”³ (*TJ*, p. 543). Since the principles of justice are defended by showing that they would be chosen in the OP, these remarks suggest that Rawls *did* rely on (1.1) or on some premise quite like it in *TJ*, and that he relied on it to defend the principles.

Furthermore, some of the crucial assumptions that underpin the Pivotal Argument—such as those made in the moves to steps (1.3), (1.6), (1.7), and (1.9)—seem to be assumptions on which Rawls relied. (1.3) expresses a quintessentially contractualist idea about what respect for persons requires. In moving from (1.5) to (1.6), the Argument assumes it is imperative to respect persons as the kind of being (1.1) says they are. This is an imperative Rawls is widely read as presupposing and, indeed, reliance on it may seem to be the source of much of his view’s appeal. The step from (1.6) to (1.7) is taken on the basis of a claim Rawls seems to make explicitly, when he says that principles which are adopted without “exploitation of the contingencies of nature and social circumstance” express respect for those who live under them (*TJ*, p. 179/156). As we shall see later, (1.9) is necessary to sustain the Kantian Interpretation of justice as fairness (cf. *TJ*, p. 252/222).

But the Pivotal Argument is not one that Rawls ever lays out systematically nor can it be extracted from any one passage of *TJ*. This may engender some doubts about the claim that Rawls relies on it or any argument like it, and so may raise doubts about whether there is any plausible reading of Rawls that gives it a central place. These doubts may be heightened by two clearly identifiable ways in which the Pivotal Argument diverges from *TJ*’s defenses of the principles of justice, for if the Pivotal Argument omits considerations or arguments on which those defenses draw, then the *Public Basis View*’s claim to identify the sources of Rawls’s dissatisfaction with those defenses would be undermined.

One especially notable and surprising departure from Rawls’s texts seems to be that the Pivotal Argument accords the OP only derivative force in support of the principles of justice: the OP is not referred to explicitly until (1.9) and the argument does not go through the details of the parties’ choice there. On the contrary, I think the secondary role given the OP tells in favor of imputing the

2. See Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1982), pp. 48ff., 133–34, and 175ff.

3. I take this passage to support the claim about freedom and equality, and not just freedom, because I take Rawls to mean that parties assume they all will have an equal interest in liberty in due course.

Pivotal Argument to Rawls rather than against it. For this reason, and because the objection raises deep issues that I shall take up later, I want to confront it.

In imputing the Pivotal Argument to Rawls, the *Public Basis View* builds on an insight that was first articulated by Ronald Dworkin. That insight is that in *TJ*, Rawls argues “through” the OP from more fundamental presuppositions.⁴ Describing how he thinks Rawls argues through the OP, Dworkin writes:

The original position is well designed to enforce the abstract right to equal concern and respect, which must be understood to be the fundamental concept of Rawls’s deep theory.⁵

The Pivotal Argument seems to spell out Dworkin’s insight by showing that Rawls argues through the OP in just this way. For at (1.10), the OP seems to do the enforcing to which Dworkin refers. The transition from (1.5) to (1.6) seems to depend on the right to respect that Dworkin says it enforces.

Dworkin’s reading of Rawls is open to question. Moreover, there remains some controversy about just what Dworkin has shown even if his interpretation is right. That controversy bears on the plausibility of the *Public Basis View* and of other views, like my own, that attribute something like the Pivotal Argument to Rawls. Dworkin is sometimes thought to have shown, not just that Rawls argues through the OP, but that the OP is therefore dispensable. If this reading of Dworkin were correct, and if Dworkin’s reading of Rawls is correct, then that would tell against imputing the Pivotal Argument to Rawls since the Pivotal Argument goes through the OP, but does not dispense with it. But this reading of Dworkin is a mistake. Dworkin argues that the OP does not have fundamental justificatory force. As I shall explain in §VII.9, nothing he says entails that it is dispensable. So Dworkin’s reading does not imply that the Pivotal Argument should not be imputed to Rawls.

Someone working in the spirit of Dworkin *could* show that the OP is a dispensable part of the argument for C_1 —the claim that primary goods must be distributed in accordance with the principles of justice—by producing an argument for C_1 that begins from the requirement of equal concern and respect but does not go by way of the OP. As we shall see in Chapter VII, such an argument in effect moves from (1.6) to C_1 differently than the Pivotal Argument does, by attaching a different interpretation to (1.6)’s requirement that principles be acceptable to us “as free and equal persons.” Dworkin himself does not provide such an argument, but Joshua Cohen does.⁶ In an important paper called “Democratic Equality,” Cohen argues for Rawls’s principles

4. Ronald Dworkin, “The Original Position,” in *Reading Rawls* (Oxford: Basil Blackwell, 1975), ed. Norman Daniels.

5. Dworkin, “The Original Position,” *Reading Rawls*, p. 181.

6. Joshua Cohen, “Democratic Equality,” *Ethics* 99 (1989): pp. 727–51. Cohen does not cite Dworkin, and I do not mean to suggest that Cohen himself accepts Dworkin’s interpretation of Rawls.