Reconsidering Rawls: The Rousseauian and Hegelian Heritage of Justice as Fairness

by

Jeffrey Neil Bercuson

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Political Science
University of Toronto

© Copyright by Jeffrey Neil Bercuson 2013
Reconsidering Rawls: The Rousseauian and Hegelian Heritage of Justice as Fairness

Jeffrey Neil Bercuson

Doctor of Philosophy

Political Science
University of Toronto

2013

Abstract

This dissertation is an attempt to better understand the moral and political thought of John Rawls. I begin by calling into question the conventional, though misleading, image of Rawls as a thoroughgoing Kantian. While the influence of Kant upon Rawls is undeniable and therefore well documented, there are important theoretical differences between them, and these differences open up the necessary interpretive space for the under-appreciated influences of Jean-Jacques Rousseau and G.W.F. Hegel. That neither Rousseau – a theorist of recognition – nor Hegel – a theorist of reconciliation – is regarded as an important influence on Rawls is a major oversight in the history of political thought – an oversight that my dissertation hopes to amend. But there is more at stake here than the addition of a new chapter in the history of political philosophy: when we expose the full extent of the Rousseauian and Hegelian heritage of justice as fairness (and later, political liberalism), we get a more complete, nuanced – and, in my view, a more attractive – image of the moral and political philosophy of Rawls. This new, richer image of Rawls’s political philosophy is captured by what I call “robust reasonableness”: what Rawls offers, in the end, is a more conspicuously demanding account of the reasonable – of our obligations towards our fellow participants in social cooperation. Justice as fairness is thus anchored by a morality of engaged and committed citizenship. This is precisely what Rawls sees as missing from Kant’s ethical philosophy. In response, he turns to Rousseau and to Hegel, both of whom provide, at least on Rawls’s view, persuasive solutions to the pathologies of social and political life. Rawls incorporates many of these solutions into the normative and practical landscape of his own philosophical doctrine, and this compels us to reconsider that doctrine in the light of these unrecognized influences.
Acknowledgments

I would like to thank the members of my committee, Nancy Bertoldi, Ronald Beiner, Joseph Carens, Simone Chambers and Anthony Laden.
Table of Contents

Introduction ......................................................................................................................... 1
Chapter 1: Beyond Kant ..................................................................................................... 9
Chapter 2: The Hegelian Dimensions of Justice as Fairness ........................................... 33
Chapter 3: The Rousseauian Dimensions of Justice as Fairness ........................................ 68
Chapter 4: Bringing Robust Reasonableness Into View .................................................. 97
Chapter 5: The Width of Public Reason ............................................................................. 118
Conclusion .......................................................................................................................... 146
Bibliography ....................................................................................................................... 158
Introduction

This dissertation is a re-evaluation of the moral and political philosophy of John Rawls. The goal is to better understand Rawls’s thought in light of his engagement with Rousseau, Kant and Hegel; specifically, the dissertation will highlight Rawls’s departures from Kant – despite Rawls’s generally accepted thoroughgoing Kantianism – and the unfamiliar, under-appreciated influences of Rousseau and Hegel on justice as fairness. I hope to show that the story about Kant is not the whole story about Rawls: when we look at Rawls through the lens Rousseau and Hegel as well, we get a deeper and more accurate picture of Rawls’s moral and political philosophy. This new, richer image of Rawls’s political philosophy is captured by what I call “robust reasonableness”: my aim in the present dissertation is to show that Rawls’s notion of reasonableness – of the kind and extent of our obligations to those fellows with whom we are engaged in social cooperation – is conspicuously more demanding, and therefore more attractive, than most interpreters and critics assume. This is the core of the original contribution offered in this dissertation: what I also hope to show is that Rawls’s political philosophy is best characterized by a quite demanding morality of citizenship and that this is precisely what he sees as missing from Kant’s ethical philosophy. And so, he turns to Rousseau and to Hegel, both of whom provide (at least on Rawls’s view) attractive images of engaged citizenship worthy of emulation. For Rawls, that is, the political philosophies of Rousseau and Hegel each offer persuasive solutions to the pathologies of social and political life, many of which he incorporates into the normative and practical landscape of justice as fairness. This helps to clarify my present ambitions: this dissertation is not primarily an intellectual history. Rather, it is an attempt to better understand Rawls’s normative thought in light of his views on the history of political thought.

Despite that specific substantive aim, in the chapters that follow there are brief excursions into the thought of Kant, Rousseau and Hegel: once we see the extent to which Rawls, Rousseau and Hegel are animated by similar concerns and offer similar solutions to social and political problems, we get a clearer image of the Rousseauian and Hegelian heritage of justice as fairness. But it must be noted, here, at the outset that the primary source for these discussions of past thinkers is Rawls himself. Indeed, the original idea for this dissertation came with the (relatively) recent publication of Rawls’s Lectures on the History of Moral Philosophy (2000) and his Lectures on the History of Political Philosophy (2007). It is my view that these texts
represent an invaluable resource for better understanding Rawls’s own thought: he often makes reference to his own normative political doctrine – justice as fairness (and, eventually, political liberalism) – and to the way in which that doctrine was fundamentally shaped by his interpretations of the canonical thinkers included in his Lectures. This raises another important point worth emphasizing at the outset: many readers will be struck by the peculiar, idiosyncratic nature of Rawls’s interpretations of Rousseau, Kant and Hegel (and of the first two in particular). Indeed, Rawls’s Lectures are often characterized by their rejection of some popular academic consensuses. This is an important methodological moment: it is not my purpose in this dissertation to arbitrate interpretive controversies about Rousseau, Kant and Hegel. We are not, in other words, concerned with the following kinds of questions: Is Rawls’s interpretation a good interpretation? Why did Rawls endorse such a peculiar or idiosyncratic or narrow image of the thinker in question? Instead, we accept Rawls’s interpretations as his own genuine, thoughtful understandings of complex texts. Once again, our ambition is not to better understand the history of political philosophy. Our purpose is to better understand Rawls’s normative thought in light of his unique engagement with some of the canonical thinkers in the history of political thought.

What, exactly, do we learn about Rawls from his moral and political Lectures? Let us begin with the question of Immanuel Kant. It is my view that Rawls’s dissatisfaction with certain elements of Kant’s thought sets the stage for the broader questions and themes illuminated by Rawls’s engagement with Rousseau and with Hegel. And so, I have made the chronologically peculiar decision to begin the dissertation with a discussion of Kant’s important, but importantly limited, influence on justice as fairness. Chapter 1 thus outlines both the Kantian dimensions of justice of fairness, as well as the crucial limitations of a strictly Kantian framework when trying to fully understand Rawls’s normative political thought. Of course, there is no denying the immense influence of Kant on Rawls: Rawls’s repeated emphasis on the “reasonableness” of liberal citizens – that is, individuals’ commitment to treating their fellows in a way that adequately respects their moral status as free and equal – and the centrality of that ideal to justice as fairness is a testament to the immensity of Rawls’s debt to Kant. Indeed, Rawls rejects the view that the categorical imperative is wholly formal – it is wrongly characterized, he says, by an exclusive focus on harm and a concomitant emphasis on negative freedom – and, by extension, accepts the view that the moral law endows agents with the obligation to actively pursue the kinds of
institutional conditions that are conducive to the effective pursuit of individual self-interest. Respect for the equal dignity and status of autonomous persons is a thoroughly Kantian ideal, and no commentator can deny in good conscience the influence of this ideal on Rawls’s thought.

But what Rawls’s lectures on Kant (and Rousseau and Hegel) make clear is that Rawls sees the demands of reasonableness as extending beyond this narrow Kantian formulation. In other words, reasonableness requires more than the obligatory, but ultimately indefinite pursuit of a just political community (conducive to the universal satisfaction of the self-interest of rational ends-setters). But what Rawls’s lectures on Kant (and Rousseau and Hegel) make clear is that Rawls sees the demands of reasonableness as extending beyond this narrow Kantian formulation. In other words, reasonableness requires more than the obligatory, but ultimately indefinite pursuit of a just political community (conducive to the universal satisfaction of the self-interest of rational ends-setters). Another way to formulate Rawls’s critique of Kant is this: that Kant’s thought is (according to Rawls’s Lectures) insufficiently (or indirectly) political – that the possibility of justice in the world depends, according to Kant, on (a priori) moral feeling, not political activity. This is where robust reasonableness starts to come into view: on this post-Kantian view, robust reasonableness is characterized by more precise, definite and conspicuously political content. Of course, it is a central claim of this dissertation that we get a more accurate, robust – and, in my view, attractive – image of Rawlsian reasonableness when we examine Rawls’s engagement with Rousseau and Hegel. Rawlsian reasonableness, in other words, has fundamentally and consciously Rousseauian and Hegelian features.

In this vein, the transition from Chapter 1 to Chapter 2 is instructive: Kant and Hegel are engaged in a direct conversation, and it is my view that important elements of Hegel’s critique of Kant inform, and recur in, Rawls’s thought. In fact, in his “Kantian Constructivism in Moral Theory,” Rawls explicitly adopts the Hegelian goal of overcoming the dualistic quality of Kant’s moral and political philosophy, and he specifically cites Hegel as an important precursor in such an endeavour. In what ways does Rawls emulate Hegel? As we shall see, the philosophical and methodological emphasis of Rawls’s thought shifts from the abstract, ahistorical orientation of A Theory of Justice – the centrepiece of which is the original (i.e. pre-political) position thought-experiment – to the contingent, history-driven approach of Political Liberalism. This idea is

1 See e.g. Dreben 2003, p. 340 for an exaggerated statement of this view: “Kant’s talk about practical reason is useless for understanding Rawls […] For me, whenever Rawls has gone slightly astray in exposition has been when he has paid too much lip service to Kant. It does not help at all” (Dreben 2003, p. 340). While I do not subscribe to this extreme view – surely, Dreben exaggerates for rhetorical effect – I do accept the view that “Kant’s talk about practical reason” is a limited framework for understanding Rawls. In fact, this is precisely the starting premise of the present dissertation.
captured by Rawls’s later methodological emphasis on reflective equilibrium (and not the original position) as the essential test for any set of political principles. Rawls’s debt to Hegel can therefore be summarised in the following way: the moral sentiments required for effective political activity are not (contra Rawls’s Kant) a priori or determined by a transcendental moral law; rather, such sentiments are generated by well-designed (i.e. principle-guided) political institutions. In other words, citizens are educated to this robust idea of reasonableness by the principles, values and practices that characterize our evolving social life. Human nature is thus characterized by its ultimate permissiveness: whereas, for (Rawls’s) Kant, the conscience is alive to its moral obligations regardless of the contingencies of social and political life, the Rousseauian-Hegelian-Rawlsian view admits that the morality of citizenship (i.e. reasonableness itself) is the outcome of historical processes. And so, Chapter 2 begins by conceptualizing Rawls’s later work as a shift in emphasis from the Kantian transcendentalism (or a priori-ism) characteristic of Theory to a mode of Hegelian reconciliation that emphasizes the individual’s commitment to their prevailing political-institutional milieu (as opposed to a future realm of ends as the object of a Kantian reasonable faith) partly by virtue of the pedagogical function of the coercive institutions to which individuals are subject.

This notion of Hegelian reconciliation is the first (of two) additional dimensions of robust reasonableness that emerges from a close reading of the moral and political Lectures. On this view, a constitutive aspect of Rawlsian reasonableness is active political labour in the service of institutional preservation and vitality. Here, Rawls will tell a conspicuously Hegelian story – in fact, Theory makes explicit reference to Hegel in precisely this vein – about the eventual coincidence of private and public goods. In other words, citizens come to regard the flourishing of the political community as an essential aspect of individual flourishing. The political community of justice as fairness therefore constitutes what Rawls calls a “real unity” – again, note the Hegelian language – characterized by publicly shared ends (i.e. institutional preservation, consolidation and enlargement) and by the desire to overcome the pervasive atomism characteristic of market society.

Chapter 3 continues the task of elaborating the more demanding obligations that, on my view, characterize robust reasonableness. This third chapter adopts an explicitly Rousseauian perspective: in it, I advance the view that justice as fairness is perhaps best understood in a kind of politics of recognition framework. Indeed, the extent to which justice as fairness is motivated
by psychological concerns – namely, the need for recognition and for self-respect – has not been fully appreciated by Rawls’s interpreters and critics. And Rawls explicitly follows Rousseau’s solution to the lack of recognition and self-respect: by institutionalizing the rights of citizenship in the basic structure. And so, the Rousseauian interpretation of reasonableness is the extension of institutionalized recognition to socio-psychologically complex beings that have complex psychological needs. As we shall see in detail throughout Chapters 2 through 5, reasonable treatment is a much more demanding ideal than (merely) ensuring social and economic independence through institutionalized redistribution: it requires the promotion of a maximally inclusive community and, by extension, making each member of that community feel like a meaningful part of a shared common project.

In Chapters 4 and 5 there is a conspicuous shift in emphasis: we shift our focus from the moral and political Lectures – as well as the Rousseauian and Hegelian features of justice as fairness outlined therein – to a reconsideration of Rawls’s own normative theory in light of its extensive, and yet under-appreciated, non-Kantian heritage. This shift in focus and in emphasis is both deliberate and necessary: we are not only interested in Rawls’s interpretation of the history of political thought; we are equally (and perhaps even more) interested in Rawls’s unique contribution to that on-going dialogue. What I ultimately hope to show in the final two chapters is the conscientious demandingness of this more elaborate conception of the reasonable. This newer, thicker, robust conception of the requirements of reasonableness incorporates both the Rousseauian notion of self-respect outlined in Chapter 3 and the Hegelian notion of reconciliation outlined in Chapter 2.

In this spirit, the goal of Chapter 4 is to fully draw out what robust reasonableness is, as well as the ways in which robust reasonableness contributes to communal vitality. Indeed, much of the liberal-communitarian debate that followed the publication of A Theory of Justice will be exposed as the result of a series of misinterpretations: Rawls does not subscribe to a view of persons as atomised or as fundamentally self-interested (despite his employment of the hypothetical social contract and his concomitant emphasis on individual rights), and it is equally clear that, all along, Rawls accepts and appreciates the significance of community. In the end, robust reasonableness is best characterized by an imperative of radical inclusiveness: the legitimate and stable community must guarantee every citizen’s equal right to participate in the political and economic affairs of the political community.
In my view, the most important and most frequently misunderstood statement of Rawls’s demanding morality of citizenship is his treatise of international relations, *The Law of Peoples*. What we get in that late text is Rawls’s most complete statement of the necessary and sufficient conditions of self-respect: (1) that one receives the right kind of recognition from one’s domestic institutional milieu and (2) that one regards one’s political community as in possession of a meaningful degree of political and economic independence. The self-respect of *Theory* and the amour-propre of *The Law of Peoples* will thus be characterized by their ultimate inseparability. It is precisely when we confront Rawls’s international political theory that the truly demanding nature of robust reasonableness comes into full view: international arrangements are reasonable only insofar as they promote the autonomy and concomitant self-respect of peoples. In other words, the demands of robust reasonableness extend beyond domestic borders: to the extent that economic-distributive-institutional arrangements *undermine* the self-sufficiency, and self-respect, of burdened peoples, well-ordered peoples are under an obligation of justice to reform these unequally burdensome arrangements.

Finally, the purpose of Chapter 5 is to raise important questions about the internal coherence of the philosophical project that characterizes Rawls’s later (domestic) work, *Political Liberalism*; in it, we are guided by the following kinds of questions: is Rawls’s later distinction between (merely) political and (more existentially robust) comprehensive doctrines philosophically coherent? Does the political-comprehensive distinction hold up to scrutiny? These (and related) questions are in my view demanded by the aforementioned discussion of the centrality of community in Rawls’s thought. Indeed, Rawls will often speak of the fulfilment of the demands of robust reasonableness (in the conduct of individual citizens as contributors to communal vitality) as the essential or fullest realization of human nature. Of course, this strikes many interpreters as a fundamental violation of the aspirational neutrality of political liberalism: why isn’t the kind of citizenship demanded by *robust* reasonableness a comprehensive view of the good? What about those citizens for whom the demands of citizenship are irreconcilable with other comprehensive commitments? In other words, we must confront the possibility that the Rousseauian and Hegelian heritage of justice as fairness leads Rawls towards the articulation of a more attractive practical philosophy of citizenship, but one that ultimately fails to satisfy the stringent philosophical demands *set by Rawls himself*, and which characterize his later philosophical enterprise.
Despite such concerns, I believe that Rawls’s project is both philosophically sound and practically attractive: Chapter 5, like the earlier chapters, is ultimately a defence of the Rawlsian political enterprise despite some of its philosophical ambiguities. And so, just as earlier chapters were devoted to the development of a more robust conception of the requirements of reasonableness, this final chapter will attempt to develop a more robust, more convincing conception of the idea of public reason. Abiding by the constraints of public reason – according to the wide view thereof, as articulated in Rawls’s late text, “The Idea of Public Reason Revisited” – will be shown to be a constitutive aspect of the morality of robust reasonableness. We are thus lead to two important conclusions: (1) that political liberalism, as described by Rawls, is in need of significant expansion (or elaboration), but (2) that those revisions ought to be thought of as internal to the idea of public reason at the centre of that doctrine. And so, the ultimate purpose of this final chapter is to show that Rawls himself has very compelling reasons to endorse the revised image of political deliberation outlined therein. Indeed, Rawls’s image of the well-ordered, stable, solidaristic political community remains a potent resource for rethinking some of the constraints on what is permissibly said in the public political sphere.

In the end, the well-ordered community of justice as fairness is best characterized by a powerful, pervasive ethos of commitment, trust, co-operation, mutual respect and even love (i.e. the legitimate amour-propre of The Law of Peoples). While liberalism may have been born in the aftermath of religious war – and it may have had a long period of instability following its birth – the doctrine itself contributed to the emergence of a political and social condition in which toleration, co-operation and mutual respect can overcome the enmity previously produced by comprehensive pluralism. In this spirit, the purpose of the Conclusion is to show that the morality of robust reasonableness – which, to reiterate, emerges out of Rawls’s engagement with Rousseau and with Hegel – is the basis of a very compelling response to one of Rawls’s most insightful and outspoken critics, G.A. Cohen. In fact, it is accurate to say that Cohen is the implicit interlocutor throughout the dissertation, and that the Conclusion hopes to make the critical conversation between him and Rawls explicit.

According to Cohen, then, there is in Rawls’s normative political thought a kind of psychological and motivational disconnect between the choices and values of ordinary citizens in their daily lives and the normative principles that guide the institutional distributive arrangements demanded by Rawlsian principles of justice. And, for Cohen, this disconnect is most
conspicuous when it comes to the question of distributive justice: the kinds of incentives embedded in Rawls’s difference principle – the talented are entitled to unequal distributive shares as long as the worst-off are made better off as a result of their labours – run counter to the spirit of the difference principle itself: those talented people, says Cohen, “would not need special incentives [to accumulate] if they were themselves unambivalently committed to [the difference principle]” (RJE, 32). Again: the incentives aspect of the difference principle – the incentive to get more rich than one’s fellows – dissociates the act of accumulation from the moral (i.e. autonomy-guaranteeing) purpose it is meant to serve.

What I hope to show in the Conclusion is that Cohen’s critique is a powerful but ultimately misleading one: that the kind of political community described by Rawls throughout his normative oeuvre is animated by an ethos of justice, and that that ethos is the by-product of the socializing capacities of principle-guided institutions. Cohen’s critique thus ignores the extent to which institutions can transform the values attitudes, motives and choices of individuals: just as toleration was initially regarded as modus vivendi but came to be seen as a moral principle, so too must we accept the possibility of the gradual adoption of the social or cooperative ethos implicit in the difference principle. Structures and choices may start out at opposite ends of the motivational spectrum, and then over time begin to coincide precisely because structures influence choices – institutions educate. Of course, this brings us back to the Rousseauian and Hegelian heritage of Rawls’s political philosophy: we ought not think, with Kant, of the sense of justice as a priori or immanent in human nature. Instead, Rawls follows Rousseau and Hegel in regarding the morality of citizenship – the morality of robust reasonableness – as the by-product of living under a well-ordered institutional milieu. The aim of the dissertation is to bring out this essential insight and to explore the normative and practical implications of this morality of robust reasonableness.
Chapter 1: Beyond Kant

“Every human being has a conscience and finds himself observed, threatened, and kept in awe by an internal judge; and this authority watching over the law in him is not something that he himself makes, but something incorporated into his being. It follows him like a shadow when he plans to escape.”

-- The Metaphysics of Morals (1797)

We begin by turning to Rawls’s extensive engagement with, and appropriation of, Kant’s moral philosophy. Of course, there is no shortage of literature on the important affinities between the two thinkers (Baumrin 1975, Darwall 1976, Pogge 1981 and Hill Jr. 2000). And Rawls himself is outspoken on the many Kantian dimensions of justice as fairness; indeed, nowhere is Kant’s influence clearer than in Rawls’s repeated emphasis on the “reasonableness” of citizens, who are committed to treating their fellows in way that respects their equal status as free and equal. We will flesh out this Kantian interpretation of reasonableness in considerable detail below. But the main task of the present chapter is to examine Rawls’s self-professed departures from Kant’s moral and political philosophy. Indeed, the central claim of the present chapter is this: that Rawls’s idiosyncratic interpretation of Kant informs these departures, and this is what compels the commentator seeking to better understand Rawls’s oeuvre to consider alternative, under-acknowledged influences – namely, Rousseau and Hegel, to whom we turn in Chapters 2 and 3. Indeed, what we shall ultimately find is that an exclusive focus on the Kantian heritage of justice as fairness leaves too much out: we get a much more nuanced, sophisticated and complete image of Rawls’s moral and political philosophy – and of reasonableness, in particular – when we import the questions and concerns of these other, under-appreciated influences.

Consider, in this vein, a brief example that foreshadows a later discussion in Chapter 3. Another aspect or dimension of reasonableness that goes beyond the exclusive, narrowly Kantian interpretation is rightly attributed to Rousseau: reasonableness is not merely understood as extending certain treatment to similarly constituted moral beings who are capable to deliberating upon, and executing, this or that plan of life. Reasonable treatment, on this narrow (Kantian)
interpretation, is best understood as securing the *institutional* conditions for the pursuit of self-interest (however widely or narrowly the notion of self-interest is construed). Instead, the Rousseauian interpretation of reasonableness is the extension of proper, humane treatment to socio-psychologically complex beings who have needs that are not directly related to the capacity for, or exercise of, autonomy, such as *self-respect* and the appropriate *recognition* of their fellows.² In this vein, reasonable treatment is a much more demanding ideal (than ensuring social and economic independence through institutionalized redistribution): it requires the promotion of a maximally inclusive community and, by extension, making each member of that community feel like a meaningful part of a common project. We will return to this very important idea in Chapters 3 and 4 in our discussion of the “proper” *amour-propre* of the well-ordered community, and of the centrality of that ideal to Rawls’s *The Law of Peoples*.

This necessary addendum to a strictly Kantian conception of reasonableness points towards a prominent theme of the present chapter: the insufficiently political character of Kant’s thought. Of course, this claim will give many readers pause: Kant *does* have a well-developed political philosophy, which grows organically out of his moral theory; any account of his thought that (for the most part) ignores his politics is (at best) radically incomplete or (at worst) simply wrong (see e.g. O’Neill 1986, p. 524). If that is in fact the case – and it is difficult for any reader to deny Kant’s deep interest in political questions – another important task for this chapter is to identify some of the important limitations of Rawls’s account of Kant. Consider this telling fact: of the ten lectures devoted to Kant in Rawls’s *Lectures on the History of Moral Philosophy*, which total 186 pages, the “political essays” receive attention in only two sections (§X.3 and §X.5). Together, these two sections total approximately three – *three!* – pages. And this political philosophical lacuna reappears in a number of Rawls’s other writings, most notably in Rawls’s “Kantian Constructivism in Moral Theory,” which is the subject of a lengthy discussion below. Our task in this chapter, then, is three-fold: (1) to outline Rawls’s idiosyncratic interpretation of Kant’s philosophy and (2) to briefly contrast that interpretation with a

---

² In a similar vein (as we shall see in the next chapter), the notion of *reconciliation* is also a constitutive dimension of Rawlsian (not Kantian) reasonableness. Indeed, with the exception of the very powerful quote from Kant with which Rawls concludes *The Law of Peoples* – “If justice perishes, then it is no longer worthwhile for men to live upon the earth” – the idea of reconciliation is generally regarded by Rawls as a Hegelian approach to both the practice of philosophy and to the concept of reasonableness. An obvious example of this is the Introduction to Rawls’s *Justice as Fairness: A Restatement*, which will be examined in detail in Chapter 2.
compelling alternative, what I shall call the anthropological reply. This second task is an important one: the truly idiosyncratic nature of Rawls’s reading comes into sharp relief when we contrast it with this alternative account. More importantly, we also get a clearer sense of how Rawls’s comparatively narrow understanding of Kant affects the development of his own political philosophy: if Rawls regards Kant’s philosophy as best characterized by its principled transcendence of the world – and if, as I hope to show, Rawls regards this as a fundamental flaw in Kant’s thought and sees his own work as an attempt to overcome this flaw – then we ought not be surprised by Rawls’s turning to Rousseau and (especially to) Hegel for inspiration in this endeavour. The third and final task of this chapter is (3) to illustrate how Rawls’s interpretation of Kant – which eventually leads to important departures from him – informs his (Rawls’s) own normative philosophy (as well as his approach to the practice of political philosophy). By emphasizing some of the “un-Kantian” features of A Theory of Justice and of Rawls’s later works, we open up new interpretive space for Rousseau and for Hegel. In doing so, to reiterate, we get a much more thorough, accurate portrait of Rawls’s political philosophy and of reasonableness in particular.

I.

In our attempt to draw out both the a priori and empirical dimensions of Kant’s thought, the following quote from the Preface of the Groundwork for the Metaphysics of Morals sets the stage nicely: “All philosophy so far as it rests on the basis of experience can be called empirical philosophy. If it sets forth its doctrine as depending entirely on a priori principles, it can be

---

3 In addition to the authors and texts discussed below (including Herman, O’Neill, Benhabib and Marwah), see O’Neill 1986, Louden 2000, Wood 1999, and Frierson 2003. Although there is no direct engagement between Rawls and the anthropological Kantians, the latter’s critique of the Lectures is easy to anticipate: Rawls’s near-exclusive focus on the Groundwork and the Metaphysics of Morals misses Kant’s concern with the empirical or subjective side of human life.

4 What we shall ultimately find is that both rival interpretations outlined below – Rawls’s “pure” view and the “anthropological” reply – do capture something important about Kant’s thought (each misses something important as well). But our main focus remains on Rawls not Kant: again, our purpose is to outline Rawls’s view of Kant’s ethics, and to demonstrate how that view affects Rawls’s own normative political philosophy. Indeed, it is precisely Rawls’s narrow or idiosyncratic view of Kant (as insufficiently political, as much too transcendental) that compels the reader to look at Rousseau and at Hegel, and to thus get a more nuanced picture of what the practice of liberal-democratic citizenship ought to be like.
called pure philosophy” (Gr Preface: 3; see also MdS 6:217). In which kind of philosophical inquiry does Kant engage? Empirical or pure? The conventional wisdom points to the latter: Kant’s ethical thought is characterized by its thoroughly abstract quality. The transcendental status of the moral law, and the a priori recognition of the obligations implied by it, affirm the image of Kant the pure philosopher, uninterested in the contingencies of social and political life. As we shall see in detail throughout this chapter, this is very much the view of Rawls, whose own moral and political philosophy is often characterized – wrongly, as I shall try to demonstrate – by its similarly abstract quality. In this context, we turn our attention to the following passage from the Groundwork, which provides compelling evidence for adherents of the pure view: “When applied to man [moral philosophy] does not borrow in the slightest from acquaintance with him (in anthropology), but gives him law a priori as a rational being” (Gr Preface: 7). To be human, for Kant, is to be in possession of a conscience: the appeal – the power – of Kant’s moral philosophy, on this account, is its belief that we are naturally, spontaneously subject to moral dilemmas. The conscience is alive, on this view, even in a vacuum (or a tyranny).

But Kant continues (immediately after the previous passage): “These laws admittedly require in addition a power of judgment sharpened by experience […] for man, affected as he is by so many inclinations, is capable of the idea of pure practical reason, but he has not so easily the power to realize the Idea in concreto in his conduct of life” (ibid.). This is the entry point for the recent turn in Kant scholarship towards an “impure” or anthropological account of Kant’s ethics. Rather than emphasize the transcendental, a priori orientation of the Groundwork, as Rawls does, the anthropological Kantians look for Kant in the world. How, for instance, do the norms, rules and practices of particular groups – my community – conduce to (or prevent) the recognition of the dictates of the moral law? Is there a place in Kant’s ethics for mediate ideas and institutions that educate agents to an awareness of what is required by the moral law? Or, returning to the pure view, does this constitute a form of heteronomy (as Kant understand the term)? We may happen to act in morally permissible ways – having been properly educated to an understanding of right and wrong – but genuinely moral action, for Kant, is motivated solely

---

5 Actually, it is not entirely clear from the Lectures that Rawls recognizes that the pure-anthropological debate even exists; for him the pure, a priori view is the only available view. See below for an account of Rawls’s (minimal) engagement with the “impure” view.
by respect for the moral law and not its arbitrary confluence with certain canons of moral meaning and judgment. If freedom is acting from the law – not merely in accordance with it – it follows, on this pure account, that an essential precondition of autonomy is our natural and immediate consciousness of the moral law as supremely authoritative for us as moral agents (see e.g. KP 5:4n and Gr Preface: 3). At issue, in other words, is whether our consciousness of (what is required by) the moral law is a process or a fact (O’Neill 1986, p. 533 and p. 536).

In the end, we shall find a considerable amount of evidence for both views: both the transcendental and the anthropological emphases are grounded in certain aspects of Kant’s thought. Indeed, in Kant’s eyes, as the earlier quotes from the Groundwork emphasize, both are equally necessary. Rawls, though, falls firmly into the former, pure camp: the strong interpretive emphasis in his Lectures on the History of Moral Philosophy, is on the a priori status of the moral law, and on the immediate, intuitive nature of moral judgment. In his other writings on Kant, however, Rawls expresses reservations over what he takes to be the wholly abstract, wholly a priori quality of Kant’s moral philosophy. Indeed, Rawls is unconvinced by the supposed natural immediacy of the moral law, and when he himself (Rawls) acknowledges the socializing or pedagogical function of group practices, values and (most importantly) institutions he sees this as a turning away from Kant. Again, Rawls’s eventual rejection of Kantian transcendentalism has important consequences for his own political philosophy: Rawls’s interpretive emphasis shifts from the abstract and a priori to a kind of empirical analysis of this or that political community – its genesis and history, its traditions and prevailing practices – and of the attachments and obligations that membership entails.

II.

I begin section with a disclaimer: it is here that we undertake our most detailed exposition of Rawls’s lectures on Kant’s moral philosophy. The views expressed are Rawls’s, and the passages highlighted in this section are cited by him as evidence for his views. References to Kant in Section II are references to Rawls’s Kant. And, as I have already indicated, those views are deeply idiosyncratic. Why? Because they only show one side of Kant’s thought: the transcendental, a priori, abstract side. Below, in Section III, we will spend a bit of time fleshing out the other side, the empirical or anthropological dimension, which Rawls’s Lectures
conspicuously neglect. In doing so, the narrowness of Rawls’s interpretation comes into full light, and we will be better equipped to understand the precise nature of Rawls’s *departures* from Kant.

According to Rawls, then, the central motif of the *Groundwork for the Metaphysics of Morals* – a natural starting point for a study of Kant’s ethics, for its aim is “to seek out and establish the supreme principle of morality,” the moral law – is its strong emphasis on the validity of ordinary, intuitive moral judgments (*Gr* Pref: 13). Indeed, throughout the *Lectures*, Rawls emphasizes the natural immediacy of conscience, as well as the non-deducible character of the moral law, in Kant’s ethics: “The power of choice is *directed immediately* by pure reason’s idea of the moral law” (Rawls 2000, p. 263 italics added). Again: “In our *common moral consciousness* we recognize and acknowledge the moral law as supremely authoritative and *immediately directive* for us” (Rawls 1999a, p. 517 italics added). In his attempt to show the *a priori* immediacy of the moral law, Rawls highlights two important passages from the second *Critique*: “The moral law is given, as an apodictically certain fact, as it were, of pure *reason*, a fact of which we are *a priori* conscious, *even if it be granted that no example could be given in which it had been followed exactly*” (*KP* 5:46ff italics added). Again: “The *a priori* thought of a possible universal law-giving [...] *without borrowing anything from experience* or any external will, is given as an unconditioned law” (*KP* 5:31 italics added; see also Reiss 1971, p. 70 – 71 and 80 – 86).

Our consciousness of the moral law, on this account, is not the outcome of moral experience or of our exposure to the justice principles that guide our shared public order. Instead, the moral law is “authenticated” by the “fact of reason” (Rawls 2000, p. 267; see also *KP* 5:16, 5:27 and 5:30). Practical reason is thus fully suited to the task of assessing the moral worth of actions. Says Kant, in this vein, in the Preface to the *Groundwork*: matters of morality are “easily brought to a high degree of accuracy and precision even in the most ordinary intelligence” (*Gr* Pref: 11). Again, Kant: “We cannot possibly conceive of reason as being consciously directed from the outside in regards to its judgments” (*Gr III*:4). “Practical reason,” echoes Rawls, “is manifest in

---

6 Rawls puts an equally strong emphasis on the following quote from the first *Critique*: “[R]eason does not [...] follow the order of things as they present themselves in appearance, but *frames for itself with perfect spontaneity* an order of its own according to ideas [of pure reason], to which it adapts empirical conditions” (*KR B576* italics added). See also Allison 1989, Beiser 1992 and Ameriks 1982.
our everyday moral thought, feeling and conduct” (Rawls 2000, p. 162). We may at times be tempted to act for the wrong reasons – the subjective disinclination to follow the moral law is a permanent fact in Kant’s ethics, and this is a rare moment of agreement between the pure and anthropological Kantians – but that does not obscure the essential truths that the *Groundwork* attempts to convey: that practical reason is fully capable of recognizing the duties and obligations associated with the moral law – *that we are immanently free regardless of socio-political circumstance* – and that we are equally capable of sophisticated moral reflection.

The *Groundwork*, then, is not an attempt to teach right and wrong: Kant must dismiss such an endeavour as “presumptuous,” says Rawls, for our duties and obligations are rendered fully intelligible to us by practical reason itself (Rawls 2000, p. 148). Instead, the *Groundwork* is Kant’s attempt to illuminate the underlying principle – or principles, if we take into account Kant’s three distinct formulations of the categorical imperative⁷ – that informs everyday (intuitive) judgments of moral worth (*Gr* I: 20; see also *KP* 5: 3). What makes such knowledge (and self-conscious reflection) possible in the domain of morality? According to the Rawlsian interpretation of Kant, the moral law is known to us *a priori* precisely because it *originates* in our free reason (Rawls 2000, p. 147; see e.g. *Gr* Pref: 6-8; ibid. II: 90). Kant is also committed to the belief, adds Rawls, that being fully conscious of our status as the source and origin of the moral law does the motivational work required to follow its dictates (ibid.; see e.g. *Gr* II: 9 and *KP* 5:74).

It follows, on the Rawlsian interpretation, that the moral law *itself* is an incentive to moral action. Human beings are characterized by their unique moral sensibility: we are naturally endowed with the capacity to understand and to intelligently apply the moral law (see e.g. *MdS* 6:213f and *Rel* 6:36). This is also another way of describing our free power of choice: we are naturally predisposed to follow the moral law *despite* its co-existence with the sensuous (or phenomenal) needs and inclinations. Indeed, doing what morality requires *despite* the inclinations is

---

⁷ In the first formulation, we examine the maxim from *our* point of view; in the second formulation, however, we are forced to consider the maxim in question from the point of view of those affected by our actions. In the third formulation – which enjoins us to construct the realm of ends – we no longer regard ourselves as simply *subject* to the moral law, but as *lawmakers* for a moral regime. The unifying idea, though, despite these multiple formulations, is that we ought to respect the moral powers – the capacity for rationality and for reasonableness (the important difference between these two concepts is explained below) – both in ourselves and in others. See, e.g. *KP* 5:151, *Gr* II: 75, *MdS* 6: 400 and Schneewind 1992 §VI.
autonomy; we are, by virtue of this capacity for choice, free moral agents and, by extension, deserving of the respect and restraint of our fellows (KP 5:47; Gr II:55 – 80; Mds 6:226). The moral law therefore has what Kant calls regulative priority over our actions: it and the inclinations are simultaneously present in our moral dispositions (Rawls 2000, p. 305).

An important implication of this pure view, again, is that our inner predisposition to moral personality is the essential foundation of autonomy (as understood by Kant): when we act in ways contrary to the moral law, we do not reproach ourselves with reference to an antecedent or external order of values (given to us by God, say) – the failure to follow the moral law leads to self-reproach (Rel I 6:36, Mds 6:380 and Reiss 1971, p. 89). According to Rawls, such a view has a crucial interpretive implication: “We are not to regard our fundamental character as determined in time; that is, we are not to regard it as a social artefact, or as the product of happenstance […] It is an evasion of our responsibility to say that we are constituted this way or that by society or by anything external to our reason and will” (Rawls 2000, p. 299; see e.g. Rel 6:35). Echoes Lilla: “If man is to become fully responsible for his actions, at every moment, he can no longer shift responsibility back to original sin or plead extenuating circumstances” (Lilla 2007, p. 149 italics added). Our fundamental character, that is, is best expressed in moral thought guided by pure practical reason, and must be regarded as wholly separate from accidental contingencies such as religious or cultural commitment or political circumstance. Says Kant, in this vein: “Through no cause in the world can [human beings] cease to be freely acting being[s]” (Rel 6:41; see also KP 5:99).9 The moral sensibility is a testament to our independence from the natural and social world. This is the guiding theme of Rawls’s Lectures on Kant’s ethical philosophy.


9 For Kant, the capacity to do so – the natural predisposition to autonomy – is a testament to God’s existence and to His desire to see “the highest good” realized (KP 5:130f).
III.

We turn now to the second of the three tasks outlined in the Introduction: to briefly outline a compelling alternative to the transcendental, *a priori* emphasis of Rawls’s *Lectures* on Kant. Why do the anthropological Kantians depart from Rawls’s reading? Here, following the work of Barbara Herman and Onora O’Neill (among others), we begin to call into question the immediacy of conscience – in Rawls’s presentation of Kant’s ethics, one brief mention aside (discussed below), moral judgments are characterized by their immediate, intuitive character – and, in its place, emphasize the pedagogical function of membership in specific political and cultural groups. According to this alternative school of thought – alternative, that is, to Rawls’s comparatively narrow view – we are *educated* to recognize the salient features of a given moral dilemma; intuition and natural feeling are, in other words, inadequate guides to morally appropriate behaviour. Says Kant in this vein: “The science of the rules of how man *ought* to behave is practical philosophy, and the science of the rules of his *actual* behaviour is anthropology; these two sciences are closely connected, and *morality cannot exist without anthropology*, for one must first know of the agent whether he is also in a position to accomplish what is required from him that he should do.” (*LE* 27:244 italics added; see also *OE*, p. 6).

Indeed, on this account of Kant’s ethics, the particular moral stricures of the group in question – its prevailing norms and mores – have the function of compelling the moral agent to see whether one’s *culturally-informed* maxims and actions are permissible under the moral law. What we are considering here is the potential reconcilability of “moral universalism” with a kind of “contextual sensitivity” (Benhabib 1988, p. 41).^{10}

The important point is this: that one cannot formulate a moral maxim without the appropriate, independently acquired prior knowledge (Herman 1985, p. 416; see also O’Neill 1990, p. 166). Indeed, formulating a moral maxim, which is the essential precondition of moral judgment, requires a refined sensitivity to the salient moral features of the situation in question (O’Neill

---

^{10} Here, the anthropological or “impure” view emerges as the more comprehensive or holistic account of Kant’s ethics: it puts the *a priori* and the contingent in an on-going dialogue (or dialectic). While the *a priori* law ought to constrain group practices – there is, in the anthropological account, room for basic moral intuitions about the appropriateness of equal treatment (whereas, in the *pure* account, there is by definition no room for the anthropological education to morality) – Kant is also open to the possibility of group practices illuminating (or obscuring) what is required by the moral law; *this* is precisely what is missing from Rawls’s account in the *Lectures*. 

Again, Rawls mentions, but doesn’t recognize the full importance of, the process whereby we become morally sensitive beings: “[Moral agents], let’s assume, […] know (or can formulate) the reasons from which they act, and they can state these reasons when appropriate” (Rawls 2000, p. 165 italics added). Says O’Neill of this questionable assumption: “To suppose that they can instantly recognize their situation as having a certain specification simplifies, indeed falsifies, the predicament agents face” (O’Neill 1990, p. 181). On this impure ethical account, then, Kant distinguishes between the objective determination of moral action – the a priori, transcendental moral law – and the subjective disposition, developed over time under unique empirical circumstances, that leads the agent in question to adopt the appropriate reasons for acting (Kleingeld 1999, p. 68). “Subjectively,” adds Marwah, “Kant is compelled to demonstrate how it is that human beings, endowed with a potential capacity for moral action, become inclined to recognize and internalize [the duties associated with the moral law]. Moral feeling is a matter of cultivation” (Marwah forthcoming, p. 20).

A maxim-formulator thus comes to perceive a world with moral features (Herman 1985, p. 419; see also O’Neill 1990, p. 180). These morally relevant features of the situation are what Herman calls the rules of moral salience – they are “the structure of moral sensitivity” (ibid., p. 419). The process of internalizing these rules is another way of describing an agent’s moral education: we learn, over time, which moral actions require additional justification and which are prima facie forbidden (LE 27:287 and 27:362). Moral judgment can thus be understood as an assessment of the unique moral weight attributed to the agent’s particular circumstances: “Is my situation (morally) exceptional?” asks the conscientious agent. “What the agent is to do,” in the end, says Herman, “is determined by a balance of weights” (Herman 1985, p. 419; see also Benhabib 1988, p. 34).

The important point, though, is this: that the ability to recognize certain aspects of a situation as morally relevant is not the necessary product of the faculty of reason, as Rawls’s account indicates. Rather, the rules of moral salience are “acquired in childhood, as part of socialization” (Herman 1985, p. 419). We are educated to be aware of, and to give adequate weight to, those morally salient features of a given situation. “Small children,” for example, “do not know the range of things that cause pain and need to be instructed that such and such a behaviour is hurtful and that what is hurtful or unnecessary must be avoided” (Herman 1985, p. 423). Membership in a community characterized by its acute and pervasive moral sensibility is therefore the essential
precondition for accepting, for example, a maxim of common interest. As we have already seen, there is no sense in Rawls’s Lectures that context-variability – that being a member of this or that cultural group or political community – would affect the orientation of our moral compass or, by extension, the outcome of our moral judgments. In fact, in his reply to Sulzer, which is emphasized by Rawls (p. 202), Kant claims that “moral instruction” is “ineffective” because “the teachers themselves do not make their concepts pure […] they try to do too well by hunting everywhere for inducements to be moral” (Gr II: 33n; see also ibid. II:25 – 30, MdS 6:483 and LE 438 - 439).11

Of course, the impure ethical view has important political implications: if we are educated to an awareness of the existence and importance of the moral law, then the primary pedagogical mechanism is exposure to particular sets of justice principles; the notion of the possible institutional protection of the rights of others – and the concomitant satisfaction of their needs – is a learned one. The impression we get from Rawls’s Lectures, by contrast, is of fully formed adults who are by definition adequately sensitive to the plight of their fellows. After all, according to Rawls’s view, the extent of the agent’s moral knowledge rests only with the duties implied by the moral law itself. And as we have already seen, the origin of the moral law in our persons lends the duties implied by it a kind of immediacy – an immediacy that marginalizes the importance of processes of moral maturation. The important point, however, is that agents cannot simply will a more refined sense of sympathy. Moral sensitivity, to reiterate, is the outcome of long historical processes.

Simply put, Rawls does not see this in Kant: as we shall see throughout the remaining chapters of the dissertation, Rawls is deeply interested in processes of moral education and maturation, but doesn’t regard this as part of the Kantian heritage of justice as fairness. Indeed, there is no

---

11 Consider, as a contrast, the following passage from On Education: “the principle of education which those men especially who form educational schemes should keep before their eyes is this – children ought to be educated, not for the present, but for a possible improved condition of man in the future; that is, in a manner which is adapted to the idea of humanity and the whole destiny of man” (OE, p. 14; see also ibid., p. 30, p. 84, p. 98 and p. 105, LE 27: 287 and 27:430, MdS 6:477 – 484 (especially §52), and KR B173 – 174). In her Constructions of Reason, O’Neill also indicates her disagreement with such view of Kant’s (anti-) pedagogical approach: “[Kant] favours the use of examples in educating the power of moral judgment […] Kant is distinctive for his articulation of the relationships between moral theory or principles, illustrative examples and the judgment that is involved in actual moral decisions, none of which he thinks a dispensable part of the moral life” (O’Neill 1990, p. 168; see also Marwah forthcoming).
indication in Rawls’s lectures on Kant of the capacity of political institutions — and the duties of justice instantiated by them — to affect our knowledge of what is required by the duties of virtue. There is no overlap, that is, between the doctrine of right and the doctrine of virtue; the latter are logically prior to the former. Says Rawls in this vein: “Our making of laws as we intelligently and conscientiously follow the principles of practical reason constitutes, or constructs, the public moral law for a realm of ends […] it is our pure practical reason that orders the social world as it introduces systematic conjunction into a whole of ends” (Rawls 2000, p. 203 and p. 215 italics added; see e.g. KP 5: 15, 16, 63 – 65 and Reiss 1971, p. 125). Again, Rawls: “Pure practical reason must construct out of itself its own object […] [The realm of ends] is developed from principles originating in our practical reason” (Rawls 2000, p. 226). In fact, the imposition of justice duties upon the domain of morality would, on Kant’s view, have the effect of rendering the agent’s actions heteronomous. The well-ordered republic, says Kant, in a similar spirit, “is derived a priori from reason” (MdS 6:338; see also ibid. 6:355 and Reiss 1971, p. 80 – 81 and p. 86). Again: “The doctrine of right needs metaphysical first principles” (ibid. 6:375).

What, exactly, does all of this mean? Clearly, Kant doesn’t believe that we are immediately and spontaneously able to adhere to the moral law. The moral law, recall, is experienced as a constraint on action: it applies to moral agents who are also potentially controlled by the desires and inclinations that are a necessary part of human phenomenality (Gr 4:413; MdS 6:405 – 409). We are part of the order of nature, and are not disembodied “moral machines.” What Kant does think, however, at least on Rawls’s presentation of his thought, is that we are naturally and

12 What is Kant’s definition of heteronomy? Whenever the object of our actions is prior to action itself, and when that object specifies the ends we ought to pursue (and the means by which we ought to pursue it), heteronomy results. See, for example, Gr II:5, II:75, II:81 and II:88 and MdS 6:382 – 383. Says Rawls: “[Kant’s definition of heteronomy] says that there neither exists nor subsists any such object, whether the Supreme Being, or a given moral order of values, or an order of nature […] Pure practical reason must be the supreme maker of its own principles” (Rawls 2000, p. 228; see also Rawls 1999a, p. 311 for a discussion of the difference between pure and perfect procedural justice, and the relevance of this distinction to justice as fairness and to the original position that instantiates that regime). And according to Guyer, Kant believed that philosophy “could never begin with definitions but only with certain primary fundamental judgments the analysis of which could lead to definitions at its conclusion, not commencement” (Guyer 1992, p. 8; see also Beiner 1983, p. 47 – 48 and 56 – 58). Finally, see Schneewind: “No authority external to ourselves is needed to constitute or inform us of the demands of morality […] not the magistrates and their laws” (Schneewind 1992, p. 310 italics added).

13 As we shall see in considerable detail in the next two chapters, this is the precise opposite of the view Rawls will endorse in Political Liberalism and in Justice as Fairness: A Restatement: morality and politics are ultimately characterized by their constitutive connection — by their co-existence and mutual enlargement in an evolving dialectic — and not by the necessary primacy of one doctrine over the other.
immediately conscious of the law; rational beings are, by definition, capable of thinking about themselves as the subject of moral questions or problems and are aware of the concomitant duties implied by the moral faculty. In this context, the following passage from Rawls’s Lectures is worth quoting at length: Kant “believes that there are certain moral dispositions – moral feeling and conscience, love of one’s neighbour and respect for oneself – that are natural dispositions of the mind […] they are antecedent dispositions on the side of feeling. Our awareness of these dispositions is not of empirical origin, but is known to us only from a knowledge of the moral law and of its effect on our sensibility” (Rawls 2000, p. 189 italics added; see also ibid. p. 299 – 301 and MdS 6:399 and 6:457).

To place oneself at the centre of a moral dilemma is, in other words, the fundamental expression of one’s humanity. The ability to do so is also, according to a religious interpretation of Kant’s ethics, the quintessential sign of God’s existence, of the world’s “full perfection as a beautiful moral whole” (MdS 6:459 italics added). Again, Rawls: “Kant believes that our everyday human understanding is implicitly aware of the requirements of reason […] the power of choice is directed immediately by pure reason’s idea of the moral law” (Rawls 1999a, p. 519). We can, that is, spontaneously represent ourselves as persons in possession of a moral personality, as the subject of intuitively resolvable moral dilemmas – as citizens of the mundus intelligibilis (Gr II:78). This is what Kant in the second Critique calls the “fact of reason.” The moral law is, in other words, a non-deducible fact – it is by virtue of the rational faculty’s existence.

Says Rawls of this crucial point: “We long to derive [the moral law] from some firm foundation, either in theoretical reason or in experience, or in the necessary conditions of a unified public order of conduct […] But none of these authentications are available within Kant’s critical philosophy” (Rawls 1999a, p. 523). Consider, in this context, the following decisive passage from the second Critique, a passage emphasized by Rawls in his Lectures: “The entire course of [the subject’s] existence as a sensible being is never to be regarded in his consciousness of his intelligible existence as anything but the consequence and never the determining ground of his causality as noumenon” (KP 5:97 – 98 italics added). The empirical world – the domain of sensibility and experience – ought not “trespass” into the domain of pure reason. Another important implication of Kant’s view is this: that no one is prevented from attaining moral worth because of the contingencies of political circumstance or deficiencies of character (Schneewind 1992, p. 327; see also Gr II: 65). Moral feeling transcends circumstance, “for conscience speaks
involuntarily and unavoidably” – *this*, to reiterate, is the central theme of Rawls’s lectures on Kant’s ethics.\(^{14}\)

As we shall see below, though, beginning with *A Theory of Justice*, Rawls repudiates this notion of the non-deducibility of the moral law and emphasizes instead the socializing – indeed, the *moralizing* – function of principle-guided public institutions. This interpretive tendency will manifest itself more clearly, and more centrally, in Rawls’s *Political Liberalism* and *The Law of Peoples*. In the end, Rawls doesn’t see Kant *in the world*, and this fundamentally informs his treatment of, and eventual departure from, him. Instead, Rawls’s strong emphasis rests on the “fact of reason” as a crucial interpretive key to Kant’s thought. Herman, by contrast, as a representative of the anthropological or impure view, builds her account of Kant’s ethics in light of a comparatively *unholy* view of the world: because the (noumenal) moral law defies embodiment in the natural (or phenomenal) world, it must be approximated in particular canons of socio-cultural meaning. The rules of moral salience are, in other words, the necessary by-product of the strict noumenal-phenomenal dualism that characterizes Kant’s thought, and that is consistently overlooked by Rawls in his *Lectures*.

IV.

We turn now to the third and final task of the present chapter: to illustrate how Rawls’s interpretation of Kant informs his (Rawls’s) own normative philosophy. Indeed, despite the initial postulations of difference in the Introduction, the affinities between Kant and Rawls are undeniable: the original position – described by Rawls as a “Kantian constructivist” device – is deeply and fundamentally indebted to Kant’s ethics. It functions as a kind of mediating device between the ideal of moral persons and the ideal of a well-ordered, just society *constructed* by their moral sentiments. There are three main respects in which this is so. First, the original position employs the Kantian conception of persons as both rational and reasonable: legislating participants are regarded as “self-authenticating sources of valid claims,” interested in securing

\(^{14}\) The anthropological reply is obvious and justified: Rawls takes the *Groundwork* to be the *whole* of Kant’s moral philosophy. See e.g. *Gr* II: 35.
the (institutional) conditions necessary for the pursuit (and revision) of their particular goods, and recognize the same capacities and interests in their fellows. This latter tendency is what Rawls refers to as the sense of justice, as the mark of reasonableness. Adds Rawls: “The terms ‘reasonable’ and ‘rational’ remind us of the fullness of Kant’s conception of practical reason and of the two forms of reason it comprehends” (Rawls 1999f, p. 504). Reasonable moral agents are characterized by their capacity and desire for mutually advantageous social co-operation – by the desire to secure the conditions of autonomy for their fellows (TA, ch. 8). Of course, this citizens’ mentality carries over to the day-to-day life of the political community.

Second, participants in the original position are fairly situated: each is in possession of equal legislating power, and each must be given the opportunity to signal their public assent to the proposed principles. Such equal treatment is a quintessential mark of a well-ordered society. And third, agents are deprived of particular information about their ends or abilities that could otherwise be useful in securing an asymmetrically advantageous social position: they are behind a veil of ignorance, which prevents their knowing their place in society, their class position, or their luck in the distribution of natural talents (TA, §4). The veil thus nullifies any arbitrary or accidental bargaining advantages.

The outcome of the original position is familiar: the most rational principles to choose emphasize the maximum liberty for the pursuit of particular goods, as well as redistributive practices whose purpose is to create the conditions necessary for the maximally inclusive exercise of individual agency (TA, §11 – 12). Like Kant, Rawls emphasizes the autonomy of original position participants: the chosen justice principles are dependent on a particular conception of the person – as well as on the fair organization of participants – not on some antecedent moral order or set of values. There is no external viewpoint from which agents reason about the institutional landscape of society; the veil of ignorance is a kind of metaphor for this approach. Contracting parties are thus motivated solely by their own “higher-order” interests in the exercise of their a priori moral powers (i.e. acting from the sense of justice and the formulation of revisable conceptions of the good). And this is an expression of their rational autonomy: principles are chosen in light of these regulative interests, and not with reference to some antecedent order of values or principles of right.
It is by choosing Rawls’s two principles of justice that agents effectively create the conditions in which these fundamental interests can be realized. These justice principles also fulfil a social role: by securing agreement from original position participants – all of whom are guided by the same fundamental interest in securing the institutional conditions for the exercise of autonomy – they elicit mutual consent for our background institutional arrangements. Such consent is a fundamental precondition for social stability and, over time, social solidarity: “The stability of a well-ordered society is not founded merely on a perceived balance of social forces the upshot of which all accept since none can do better for themselves. To the contrary, citizens affirm their existing institutions in part because they reasonably believe them to satisfy their public and effective conceptions of justice” (Rawls 1999a, p. 324). When the principles underlying those institutions are both fully public and universally accepted (as compatible with moral conception of the person as both rational and reasonable), “the ties of civic friendship” are thereby secured, and the stability of our political association strengthened.

But these similarities – Rawls’s extensive use of Kant’s moral philosophy as Rawls understood it – ought not obscure the main claim of this section (and of the present chapter more generally): that Rawls’s departures from Kant are equally instructive – that is, they give us a fuller image of Rawls’s political philosophy – as his appropriation of Kant’s thought. Indeed, the purpose of “Kantian Constructivism in Moral Theory” is both to indicate the influence of Kant’s thought on Rawls’s normative theory and to indicate some of the important – indeed, fundamental – differences between the two thinkers. Says Rawls: “In elaborating his moral theory along somewhat Hegelian lines, Dewey opposes Kant, sometimes quite explicitly, and often at the same places at which justice as fairness also departs from Kant. Thus there are a number of important affinities between justice as fairness and Dewey’s moral theory which are explained by the common aim of overcoming the dualisms in Kant’s doctrine,” however constitutive of Kant’s thought those dualisms are (Rawls 1999a, p. 304 italics added). Again: “Justice as fairness is not, plainly, Kant’s view; it departs from his text at many points” (ibid.).\(^\text{15}\) What are the elements of Rawls’s partial rejection of Kant?

\(^\text{15}\) See also Freeman 2003, p. 2, which is of particular relevance for us in light of the claims of the present chapter: “Rawls rejects Kant’s dualisms; he does not suppose principles of justice are a priori or based in ‘pure practical reason’ alone. Human nature and fixed empirical conditions within which practical reason is normally exercised are
According to Rawls’s “Kantian Constructivism in Moral Theory,” the fundamental flaw of Kant’s moral philosophy is its relative insensitivity to the political sphere: the realm of ends is not conceived as the product of publicly shared political principles, so the Rawlsian argument goes, but as the outcome of individual agents’ conscious adherence to a priori laws. What we are describing here is the strict bifurcation of the duties of justice and those of virtue: individual moral action effectively trumps the value and efficacy of political activity (or action through political channels). Says Rawls in this vein: “We act as individuals, and not as one corporate body, all of us acting collectively through institutional procedures” (Rawls 2000, p. 205). Kant’s ethics—rather, Rawls’s account of Kant’s ethics—is therefore geared towards the individual actor, in isolation from his or her political relations with others; on such a view, a moral world is possible if individuals fulfil their moral duties. Adds Pogge: “It is not unfair to attribute to Kant the rather naïve hope that if only everyone could be brought to make individual practical decisions in terms of the categorical imperative, the problem of social justice would take care of itself” (Pogge 1981, p. 58; see also Moore 1992, p. 55).

While the categorical imperative compels us to strive for the ethical commonwealth, politics remains, for (Rawls’s) Kant, a realm of instrumentality, necessarily subordinate to the realm of individual morality (Reiss 1971, p. 41; see also Beiner 1983, p. 66, and Benhabib 1988, p. 47). Relevant to discovering and justifying principles of justice. Rawls here moves some way toward the more ‘sentimental’ and ‘naturalistic’ account suggested by Rousseau.”

As we shall see in later chapters, such a view has important implications for Kant’s politics: we ought not conceptualize political activity, on Kant’s account, with reference to ideas about the telos or purpose of our community. This is, recall, Kant’s definition of heteronomy. Nor is there a unique class of citizens with the specialized knowledge needed to bring about such ends were they valid (Beiner 1983, p. 65). Instead, political institutions must be justified only with reference to the universal dictates of the moral law (which are, recall, perfectly clear to all rational beings) (Reiss 1971, p. 80 and p. 89). We need not be educated, on Kant’s account, to an idea of human dignity (as represented in the institutions, aims and purposes of our community), for such a notion is immediately clear via intuition. In this way, “a politics of dignity opposes itself to a politics of purpose” (Beiner 1983, p. 71). Of course, this raises an obvious problem: the moral sentiments are rarely fully instantiated in political life—history, admits Kant, is the history of human greed, vanity and selfishness despite the immediacy of the moral law—and so Kant instead relies on the “mechanisms of nature”—or, the divine telos immanent in Nature—to bring about moral outcomes (Reiss 1971, p. 41 and p. 112–113; see also Beiner 1983, p. 68 and Hedrick 2008, p. 251).

Says Doyle: “Understanding history requires an epistemological foundation, for without a teleology, such as the promise of perpetual peace, the complexity of world history would overwhelm human understanding […] Our guarantee of perpetual peace does not rest on ethical conduct” (Doyle 1996, p. 188). This rejection of a “politics of purpose” is an important moment in our attempt to better understand the intellectual relationship between Kant and Rawls. For instance, amour-propre, as described by Rawls in The Law of Peoples, describes the historical
Rawls, on the other hand, is working in distinctly social and political terms. Whereas Kant is primarily concerned with moral individuals, Rawls is concerned with moral institutions: “Justice as fairness assigns a certain primacy to the social. By contrast, Kant’s account of the Categorical Imperative applies to the personal maxims of sincere individuals in everyday life” (*PL*, p. 339; see also Rawls 2000, p. 204 and Nili 2010). This is an important point worth emphasizing: there exists, in Rawls’s thought, a set of political capacities and interests that go beyond Kant’s relatively narrow concern with individual morality. And so, we need a fuller account of what it means to act in the political community, as a member of this milieu. According to Pogge, his represents “a radicalization of Kant’s own demand for unity [of reason], now not concerned only with my but our moral fixed points (Pogge 1981, p. 62).

Although Kant and Rawls agree that our autonomy is most fully realised when we abide by laws of our own choosing, they diverge with regards to the forum in which our autonomy is most appropriately expressed. For Kant it is in our private lives – though the eventual outcome is explicitly social, the realm of ends – while Rawls insists that the primary expression of our autonomy must occur at the level of politics, by choosing and acting upon principles for institutions. Again, Rawls: “Kant proceeds from the particular, even personal case of everyday life; he assumed that this process carried out correctly would eventually yield a coherent and sufficiently complete system of principles, including principles of social justice. Justice as fairness moves in quite the reverse fashion” (Rawls 1999a, p. 339). It is of crucial importance, in Rawls’s account, that public justice comes first. And, as we shall see (in Chapter 2), Rawls will

---

17 Schneewind offers a similar account of Kant: “Kant thought that individual decision-making would be able to guide people to coordinated action on matter of general concern” (Schneewind 1992, p. 325).

18 In the Conclusion below we begin to raise questions about Rawls’s emphasis on political activity as the quintessential expression of autonomy. After all, his later thought is ultimately characterized by its rejection of such comprehensive images of human flourishing.
ultimately justify the primacy of the political with reference to the educative function fulfilled by principle-guided institutions; that is, we begin by focussing on institutions because of the educative and socializing purposes fulfilled by them.

In addition to emphasizing the distinctly political character of justice as fairness – as opposed to what Rawls regards as the comparatively apolitical character of Kant’s thought – Rawls further distinguishes his own thought from Kant’s with reference to what he calls the “full publicity” test that any theory of justice must pass; that is, in order for justice as fairness to be fully public (as it must be) it must satisfy three (publicity) conditions: first, all citizens must know and publicly affirm the principles of justice that regulate the basic structure of society; second, chosen justice principles must be supported by publicly shared methods of inquiry and of reasoning, such as the procedures and conclusions of science; and third – this is where Rawls distinguishes his own thought from Kant’s – the public conception of justice adopted must pass the test of philosophical reflection: it must be reconcilable with the considered reflective judgments of “you and me” – here and now – as the inheritors of an evolving public culture characterized by a particular set of political and legal institutions and by a particular “philosophical and historical tradition of their interpretation” (ibid., p. 324). In other words, “the full justification includes connecting [justice as fairness, as described in A Theory of Justice, which Rawls himself describes as conspicuously abstract and formal] with [a given] society’s particular conception of the person and of social cooperation” (ibid., p. 325).

Contracting parties must therefore have recourse to “publicly recognized […] sufficient reasons,” which are “rooted in the more fundamental notions of our political life,” when attempting to justify this or that set of institutional justice principles (ibid., p. 305 italics added). Indeed, on Rawls’s view, conditions for justifying a particular conception of justice hold “only when a basis is [already] established for political reasoning and understanding within a public culture […] What justifies a conception of justice is its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions [of] our public life, it is the most reasonable doctrine for us” (ibid., p. 307; see also Rawls 1999e, p.
This is precisely what Rawls calls the test of “general and wide equilibrium”: is the proposed theory of justice reasonable “given the state of public knowledge in our society” (ibid., p. 322 italics added)? After all, “there is no a priori guarantee that we have matters right” (Rawls 1999b, p. 30).

This third (strenuous) level of publicity represents an important methodological departure from Kant: whereas Kant – rather, Rawls’s interpretation of him – “thinks that the conception of ourselves as fully autonomous is implicit in individual moral consciousness” – by virtue of the fact of reason, and so “the social conditions for its realization are not emphasized or made part of the doctrine itself” – justice as fairness rests on the belief that political institutions have “decisive long-term social effects and importantly shape the character and aims of the members of society, the kinds of persons they are and want to be […] the grounds and tendency of these institutions should [therefore] stand up to public scrutiny” (ibid., p. 326). On Rawls’s view, then, the publicity condition is not merely extended to our fellows as part of the demands of equality (i.e. equal treatment requires subjecting our fellows to actions and institutions to which all can subscribe); this dimension is captured by Rawls’s employment of the Kantian adjective in his description of the original position. But, in addition, the principles of justice must be fully public because of their pedagogical or socializing function: the principles of justice “invoke a certain conception of the person. In this way citizens are made aware of and educated to this conception. They are presented with a way of regarding themselves that otherwise they would never have been able to entertain” (ibid., p. 340; see also Rawls 1999e, p. 398).

The full publicity condition is thus related to the “wide, as opposed to the narrow, view of the social role of morality [the narrow view is implicitly attributed to Kant]. The educative role of [prevailing] moral conception[s] defines the wide view” (ibid.; see also Mackie 1977, p. 106-107 and p. 134 – 136, Baier 1989, p. 774 and Raz 1990, p. 6 – 7). To put the matter differently: whereas Kant – again, Rawls’s Kant – connects the “deeper bases of agreement” on justice

---

19 And so, “the aim of political philosophy is to articulate and to make explicit those shared notions and principles thought to be already latent in [our shared public culture]” (ibid., p. 306). We take up this important question in detail in Chapter 2.

20 See also Rawls 1999f, p. 429: “We can present a political view either by starting from within a general and comprehensive doctrine [i.e. Kant’s], or we can start from fundamental ideas regarded as latent in the public culture. These two ways of proceeding are very different, and this difference is significant.”
principles to natural, immediate moral feelings – to the fact of reason – Rawls treats shared convictions about justice as a part of “the historical tradition”: “In addressing the public culture of a democratic society, Kantian constructivism hopes to invoke a conception of the person implicitly affirmed in that culture” (ibid., p. 306). Kantian constructivism – as opposed to pure Kantian ethics – looks backwards to history for necessary philosophical traction: “We are not trying to find a conception of justice suitable for all societies regardless of their particular social or historical circumstances” (ibid., p. 306; see also Rawls 1999f, p. 423). Again: “There is no alternative to founding a conception of justice suitable for a well-ordered democratic society on but a part of the truth, and not the whole, or, more specifically, on our present commonly based and shared beliefs” (ibid., p. 329 italics added). And so: “Justice as fairness seeks to identify the kernel of an overlapping consensus, that is, the shared intuitive ideas which turn out to be sufficient to underwrite a just constitutional regime. This is the most we can expect, nor do we need to do more” (Rawls 1999e, p. 410).

This leads us to a striking new account of what the original position is: it is wrongly regarded as the search for true justice principles or as the institutionalization of a priori principles knowable via intuition; justice as fairness is not a metaphysical doctrine in this sense. Rather, it is a kind of survey of competing accounts of justice: which theory of justice is most likely to elicit the unanimous consent of contracting parties looking to set up fair, mutually advantageous terms of social cooperation (ibid., p. 399)? Of course, the range of options from which we can choose is dictated by the historical and philosophical evolution of our political community: recall that in order to be made fully public – and, by extension, in order to be acceptable to contracting parties – the principles of justice themselves must be recognizable as principles for institutions. This is the third level of publicity that goes beyond Kant’s concerns (at least on Rawls’s apolitical, ahistorical view of Kant’s ethics). “There is no other way to proceed,” says Rawls, “since we must start from where we are” (Rawls 1999a, p. 322). And in a similar vein: “We collect such settled convictions as the belief in religious toleration the rejection of slavery and try to organize the basic ideas and principles implicit in these convictions into a coherent conception of justice […] we look to our public political culture itself as the shared fund of implicitly recognized basic ideas and principles […] the original position [thus] serves as a means of public reflection and self-clarification” (Rawls 1999e, p. 393 and p. 402).
In the end, Rawls’s description of the original position as a *Kantian* constructivist device strikes the careful reader as a misnomer: although the “model-conceptions” of the citizen and of the well-ordered society are inspired by Kant’s moral philosophy, Rawls’s clearly regards these conceptions as the outcome of historical experience. The reasonable and the rational – as well as the ideal of the just, well-ordered community – are thus tinged with a kind of contingent, arbitrary character. While the anthropological Kantians are able to account for the gradual evolution of the moral sentiments – of the *education to morality* – Rawls’s *Lectures on Kant’s ethics* are conspicuously *ahistorical*. Moral consciousness, on Rawls’s account of Kant, is not contingent on the agent’s socio-political (historical) circumstance; indeed, on Rawls’s view of Kant, the conscience is alive *even in a tyranny*. This is precisely the view rejected by Rawls in “Kantian Constructivism in Moral Theory;” hence, our reservations over the title of the work: it only tells *a part* of the truth about justice as fairness. Of course, this leads us to two fundamentally important questions: (1) What does this (methodological) departure from Kant tell us about Rawls’s normative political philosophy? Why is this departure important? We take up these important questions in the Conclusion below. And (2) why does Rawls emphasize the Kantian orientation of justice as fairness when his own philosophical methodology departs so radically from Kant’s? This raises important questions about Rawls’s self-understanding and his self-description. We take up this second question in later chapters.

**Conclusion.**

We have begun to form an image of Rawls as an *historical* thinker, who sees his primary task – indeed, who sees the *primary task of political philosophy* – as the provision of a philosophical account – a *synthesis* – of the principles and institutions that have characterized the evolution of this or that (liberal) democratic community. In other words, we get an image of Rawls as primarily interested in the day-to-day workings of political life (and, by extension as less interested in the abstract formulation of principles). The purpose of this chapter can therefore be summarized in the following way: justice as fairness (as Rawls came to understand that doctrine or, rather, as he came to emphasize certain methodological aspects of that doctrine) represents the rejection of the transcendental and *a priori*, and the embrace of the historical and the contingent. And all of this is nicely captured by Rawls’s eventual philosophical and
methodological departure from Kant: we learn from the world by embracing it, and not by abstracting away from it. Rawls sees Kant compelling us to do the latter, and this is what explains his hesitations over the wholesale embrace of Kant’s thought. Again, this is related to the fundamental task of philosophy, which is to bring to the surface the latent wisdom in our history and in our evolving institutions and political practices; this is the central theme of Chapter 2. Before doing so, however, we need to say more about the meaning and importance of this departure.

There are, I think, two important points worth re-emphasizing: (1) Rawls’s rejection of certain aspects of Kant’s thought leaves us with conspicuous interpretive gaps. Indeed, when we look exclusively at the Kantian character of justice as fairness, we only get a partial, incomplete image of reasonableness: as we shall see in Chapter 3, reasonableness has an important self-respect/recognition dimension, which on Rawls’s understanding is distinct from its autonomy-guaranteeing dimension; in addition, the notion of reconciliation (to this set of institutions, and not to the telos of Nature) is also a constitutive aspect of what it means to act reasonably. This is the theme of Chapter 2 below. The important point is this: that acting in the political community – rather than contemplating the temporally distant political implications of moral action – requires certain commitments and capacities that go beyond a narrowly Kantian rendition of the sense of justice. In other words, the formal desire for autonomy – and for mutually advantageous cooperation conducive to the autonomy of all – is only one part of the demands of constructive citizenship.

And (2) reasonableness ought not be understood as an a priori moral power. Instead, it is understood as the outcome of moral experience under favourable political circumstances. This is another way of describing the pedagogical or socializing function of principle-guided institutions: whatever their genesis, principle-guided institutions lead us to an idea of ourselves and of our fellows as worthy of certain kinds of treatment. And so, learning from this set of principle-guided institutions – and recognizing the kind of protection it provides – usually implies a deep commitment to that milieu. As we shall see in Chapters 2 and 3, Rawls does not shy away from the public virtue required to act on such commitments: justice as fairness does achieve the good of community (in a way that Rawls regards as fundamentally Rousseauian, as inspired by Rousseau’s notion of proper amour-propre). But, again, this leads to a new set of concerns: does Rawls’s emphasis on the reconcilability of classical republicanism and justice as
fairness – both of which emphasize public or political activity as the essential image of human flourishing – compromise his later emphasis on the political (not comprehensive) character of justice as fairness? Indeed, we are compelled to return to this question in light of the present chapter’s emphasis on the necessarily political expression of autonomy in Rawls’s thought (contra Kant’s, which promotes a comprehensive image of autonomy at the level of individual moral activity). Does Rawls’s “descent” into the daily life of the political community – after all, he openly takes his normative cue from the conventional (although potentially latent) wisdom of this or that democratic political community – risk violating the political-comprehensive distinction that is so central to Political Liberalism?
Chapter 2: The Hegelian Dimensions of Justice as Fairness

“…so philosophy too is its own time apprehended in thoughts.”

— Hegel, *Philosophy of Right* (1821)

“When a father asked him for advice about the best way of educating his son on ethical matters, a Pythagorean replied, ‘Make him the citizen of a state with good laws.’”

— Hegel, *Philosophy of Right* (1821)

We turn now to Hegel’s influence on the moral and political philosophy of Rawls. *Why Hegel?* To begin to answer this question recall our brief mention of John Dewey in the previous chapter: “In elaborating his moral theory along somewhat Hegelian lines, Dewey opposes Kant, sometimes quite explicitly, and often at the same places at which justice as fairness also departs from Kant. Thus there are a number of important affinities between justice as fairness and Dewey’s moral theory which are explained by the common aim of *overcoming the dualisms in Kant’s doctrine*” (Rawls 1999a, p. 304 italics added; see also Rorty 1994, p. 184 and p. 188n).

There are, I think, three main interpretive issues raised by this passage, each of which is the subject of a lengthy discussion below. First, in Chapter 1, we began to open up the necessary space for a (partly) non-Kantian interpretation of Rawls: because *on Rawls’s view of Kant* the duties of virtue are *a priori* knowable, the imposition of the duties of justice into the domain of morality constitutes a form of heteronomy. The well-ordered republic, says Kant, in this vein, “is derived *a priori* from reason” (*MdS* 6:338).

Of course, Rawls rejects this approach: he is openly troubled by the transcendental, *a priori* orientation of Kant’s ethical and (by extension) political philosophy. On Rawls’s view, which is explicitly regarded by him as a departure from Kant, knowledge of our duties (as citizens) is necessarily the product of *experience in the well-ordered community*. And so, by the time of *Political Liberalism*, Rawls shifts his emphasis from ideal moral agents fairly organized in an ideal, original (pre-political) setting to the historical evolution – both institutional and ideational – of the public culture of liberal democracy. Indeed, Rawls’s emphasis will come to rest more on the political-institutional consequences of European history and less so on the outcome of his
famous (and famously abstract) thought experiment. In other words, we must avoid falling into a familiar trap in Rawls scholarship: treating the original position as historical fact – as a kind of founding moment – and not as a (mere) hypothetical thought experiment designed to represent, or “model,” our intuitions about justice as they happen to have evolved over time (see e.g. Sen 2009, p. 56; see also Rawls 2001, §6.3 – §6.5 and p. 86n, McClennen 1989, p. 9 – 12 and p. 22 and Rorty 1994, p .184 – 187). Here is a typical statement of Rawls’s methodology: “The conception of the person [as having the two moral powers] is worked up from the way citizens are regarded in the public political culture of a democratic society, in its basic political texts (constitutions and declarations of human rights), and in the historical tradition of the interpretation of those texts […] That there are such ideas in the public culture is taken as a fact about democratic societies” (Rawls 2001, §7.2 and §9.1; see also ibid., §8.2, §11.3 and §23.4).

All of this is importantly connected to the third thematic emphasis of the previous chapter: as we have seen, Rawls consistently makes reference to the pedagogical or socializing functions of institutions (and thus rejects the a priori view of moral and political duties) and to the social dimension of human nature and of the moral sentiments. This is why Rawls’s philosophical approach begins by focussing on the basic institutional structure of society. Consider, in this light, the following admission in the moral Lectures: “A Theory of Justice follows Hegel when it takes the basic structure of society as the first subject of justice” (Rawls 2000, p. 366). Why is the basic structure so important to both Hegel and Rawls? Recall the following crucial passage cited in Chapter 1: political institutions (for Rawls as for Hegel) have “decisive long-term social effects and importantly shape the character and aims of the members of society, the kinds of persons they are and want to be” (Rawls 1999a, p. 326). In the end, then, a central claim of this chapter is that (1) we get a more complete account of Rawls’s views on the pedagogical function of political institutions when we trace the Hegelian influence in his thought (Sections III – VI). Indeed, what we shall see below is that Rawls tells a conspicuously Hegelian story about the emergence in three stages and eventual flourishing of the ethical life of a given political community with specific attention paid to that community’s institutional milieu, as well as its effects on the character and aims of affected citizens. Rawls and Hegel share the foundational belief that institutions (and their guiding principles) determine the kind of people we become. Citizenship is the outcome of an educative process.
This emphasis on institutional evolution and socialization leads to the second thematic concern of the present chapter, what I shall call (2) reasonableness and reconciliation (Section VII and Conclusion). The central claim of this section is that there is, in Rawls’s thought, a conspicuous shift in emphasis, or what Rawls himself calls a “recasting” of justice as fairness: a shift from Kantian constructivism to Hegelian reconciliation (Rawls 2001, p. xv and p. 3; see also Schwarzenbach 1991, p. 543 and Freeman 2003, p. 30). On this view, reasonableness is, in part, a matter of working towards the future instantiation of the realm of ends (however unsure its complete realization may be); but it is also a matter of honouring – of consolidating – the immanent justice principles that characterize our prevailing institutions. Indeed, on this Hegelian interpretation, the process of reconciliation to this institutional milieu – as opposed to a general, Kantian sense of reconciliation to a world hospitable to the gradual proliferation of reason and of rational institutions (see e.g. Rawls 1999b, §1.1.1) – is a constitutive dimension of Rawlsian reasonableness; that is, treating others as ends requires active political labour in the service of this set of prevailing institutions – their preservation, reconstruction and enlargement (whereas Kant’s political philosophy, on Rawls’s view, demands allegiance to the forthcoming realm of ends, to a regulative Idea of reason).

Ultimately, justice as fairness will come to specify what Rawls calls a common “public political life”: in addition to the private or personal goal of individual happiness (or needs satisfaction), citizens come to regard the common life of the community – the thriving of that community – as a constitutive, necessary aspect of individual fulfilment (Rawls 2000, p. 365). The political community of justice as fairness, in other words, constitutes a “real unity,” characterized by “publicly shared ends” (i.e. institutional preservation and consolidation) and by the desire to overcome the pervasive atomism of market society (ibid.). This latter goal is directly attributed by Rawls not to Kant but to Hegel (ibid.; see also TJ, §79n). Another central purpose of the present chapter, then, is to outline the extent of Rawls’s attempted emulation of certain fundamental aspects the Hegelian political community, and (as in Chapter 1) to flag for later consideration some of the potential philosophical issues associated with Rawls’s emphasis on the intrinsic “good” of the community of justice as fairness.

Finally, this chapter will also try to present (3) a new image of the role and purpose of political philosophy as practiced by Rawls (Sections I and II below). On such a view, political philosophy itself is an exercise in civic education: part of the task of philosophy is telling the
institutional history of this or that political community. This is, I shall argue, connected to another important Hegelian dimension of Rawls’s thought – to his unique philosophical methodology, which we began to distinguish from Kant’s in the previous chapter. That is, in trying to show the historical genesis of the moral sentiments, philosophy itself has a reconciliatory function: we become attached to our institutions because we come to recognize their freedom-guaranteeing character (morally- and institutionally-speaking). “[For] Hegel and Rawls,” adds Schwarzenbach, “moral philosophy is the attempt to clarify and systematize what we have all along been doing” (Schwarzenbach 1991, p. 544). The lesson Rawls takes from Hegel, to reiterate, is that the practice of philosophy has an essential role to play in this process of reconciliation (which, in turn, is essential to the stability of the political community). Like political activity, political philosophy is therefore rightly understood as a form of civic virtue. Says Rawls in this vein: “Political philosophy may contribute to how a people think of their political and social institutions as a whole, and their basic aims and purposes as a society with a history – a nation – as opposed to their aims and purposes as individuals” (Rawls 2001, p. 2).

I.

Let us begin by returning to Rawls’s “Kantian Constructivism in Moral Theory.” Rawls begins that work with a discussion of the role of political philosophy. How, he asks, can citizens settle on a public conception of justice – one that is acceptable to all in light of shared ideas about personhood, agency, citizenship and social cooperation? From A Theory of Justice, we know that the two principles of justice receive universal assent from contracting parties. This is the

---

21 In many respects, this chapter follows Schwarzenbach’s important article, “Rawls, Hegel and Communitarianism.” This is particularly true of the methodological emphasis of Sections I and II. However, the main claim of Schwarzenbach’s analysis is that the Hegelian influence on Rawls in “indirect” – that Hegel influences Rawls primarily via the Hegelian-influenced ethics of F.H. Bradley and John Dewey (Schwarzenbach 1991, 565n). I disagree. Rawls is a close reader of Hegel – not just the British Idealists or the American pragmatists influenced by him – and this view is clearly validated by the recent publication of the Lectures on the History of Moral Philosophy in which Hegel occupies a prominent place. There is another conspicuous limitation to Schwarzenbach’s analysis: its failure to discuss in detail Part III of Theory, where Rawls discusses the gradual emergence of the moral sense necessary for citizenship. As I shall argue in detail below, this is perhaps the clearest instance of Hegel’s influence on Rawls’s moral and political thought: there is, I believe, a direct parallel between the three moments in the emergence of ethical life in Hegel and of the sense of justice (or “the morality of principles”) in Rawls.
result of the mediating role of the original position – it functions as a “bridge” between the ideal conceptions of moral agents and of the well-ordered society constructed by their moral sentiments – and of the veil of ignorance that characterizes that thought experiment (TJ, § 4). The starting premise of this chapter, however, is Rawls’s self-professed departure from the abstract methodological orientation of A Theory of Justice; the later Rawls can be understood as an attempt to historicize liberalism – to analyse liberalism’s historical genesis (although part of our inquiry, here, is also to show the ultimate continuity between these “two Rawls”). In taking up Political Liberalism, then, we follow Rawls’s proclamation from the Introduction to that work: “The historical origin of political liberalism is the Reformation and its aftermath […] Something like the modern understanding of liberty of conscience and freedom of thought began then. As Hegel saw, pluralism made religious liberty possible” (PL, Introduction p. xxiv).22

Unlike Kant, for whom our guiding values and duties are known to us a priori – at least on the view of Kant presented in Rawls’s Lectures – Rawls claims that the basis of our understanding of freedom and equality is to be found in the past and prevailing public culture of our shared democratic society: recall, from Chapter 1, that “we are not trying to find a conception of justice suitable for all societies regardless of their particular or historical circumstances” (Rawls 1999a, p. 305). Instead, Rawls is focussed on democratic societies characterized by significant pluralism. We are not engaged in the attempt to seek out some antecedent moral order – to be instantiated in political life – under the guidance of those in possession of some kind of specialized knowledge, whether moral, religious, philosophical or political. “There are,” says Rawls, in this vein, “no facts” (Rawls 1999a, p. 307). Instead, on the historical view of Political Liberalism, we can institutionalize reasonable principles in the basic structure because (over time) we have already found ourselves to be eminently reasonable – capable, that is, of extending the same liberty of conscience (for example) to our fellows that they have extended to us (see e.g. Rawls 2001, §59.2). For Rawls, in other words, the reasonable is not an a priori moral power, as it was for Kant (at least on Rawls’s reading of Kant). Indeed, when Rawls

22 Rawls emphasizes the following passage from Philosophy of Right, §270, in his Lectures: “Far from it being, or ever having been, a misfortune for the state if the Church is divided, it is through this division alone that the state has been able to fulfil its destiny as self-conscious rationality and ethical life.” Indeed, the Introduction to Political Liberalism makes clear that Rawls subscribes to some form of the “cunning of reason” doctrine that characterizes Hegel’s own thought. See also Schwarzenbach 1991, p. 547 and p. 563.
acknowledges the socializing or pedagogical function of group practices, values and (most importantly) institutions he sees this as a *turning away from Kant*. Instead, reasonableness is a way of describing the gradual adoption by citizens of the emergent, evolving principles embedded in our institutional milieu.

*This* is where political philosophy enters, and its role is so fundamentally Hegelian that one is stunned by the general dearth of analyses on Hegelian themes in Rawls’s political philosophy.²³ The following passage is therefore worth quoting at length: the role of political philosophy, according to Rawls, is to “articulate and to make explicit those shared notions and principles thought to be already latent in common sense […] [philosophy] must discover and formulate the deeper bases of agreement which one hopes are embedded in common sense, or even to originate and fashion starting points for common understanding by expressing in a new form the convictions found in the historical tradition by connecting them with a wide range of people’s considered convictions” (Rawls 1999a, p. 306; see also *PL*, §II.4.1 and §II.8.2, Rawls 1999b, §1.1.1 and Rawls 2001, §2.1 and §40.3). Again, from the Introduction to Rawls’s *Lectures on the History of Political Philosophy*: “Liberal political philosophy […] is not to be seen as a theory, so to speak. Political philosophy has no special access to fundamental truths about justice and the common good. Its merit is that by study and reflection it may elaborate deeper and more instructive conceptions of basic political ideas that help us to clarify our judgments about the institutions and policies of a democratic regime” (Rawls 2007, p. 1).²⁴ Rawls regards such a view as anti-Platonic: philosophy ought not judge (or attempt to control) prevailing practice with reference to some distant, potentially unrealizable ideal. Instead, philosophy belongs to the “general background culture” of a democratic society: it is engaged with, and reformulates in illuminating ways, the “basic ideas” embedded in our history and our practice (ibid.).

---


²⁴ See also “Justice as Fairness: Political not Metaphysical”: “Justice as fairness deliberately stays of the surface, philosophically speaking […] Philosophy as the search for truth about an independent metaphysical and moral order cannot, I believe, provide a workable and shared basis for a political conception of justice in a democratic society” (Rawls 1999a, p. 394 – 395).
In other words, political philosophy must base its arguments on ideas, concepts and values *implicit* in prevailing political institutions and social practices; philosophy must “propose certain conceptions and principles congenial to [citizens’] most essential convictions and historical traditions” (Rawls 1999a, p. 306). We can only accept those conceptions of justice with which we have grown up, whether we realize it or not; and if we do not, it is the task of philosophy to expose common sense as something importantly determined by political institutional circumstances and by the resultant social relations. As we saw in the last chapter, Rawls’s treatment of Kant is best characterized by the latter’s attempted transcendence of the social and political world: Rawls, recall, doesn’t see Kant *in the world*, and so he turns to Hegel, for whom an acute sensitivity to past and prevailing socio-political circumstance is the essential precondition for the practice of philosophy (Rawls 2000, p. 334).

II.

Whereas Kant’s idea of freedom is, *according to Rawls*, transcendental, *a priori* and non-deducible – and therefore inappropriately reduced to one’s prevailing political milieu – Hegel’s view of freedom is “distinctively *institutional*” (Rawls 2000, p. 330). This connects, says Rawls, “with [Hegel’s] view of persons as *rooted in and fashioned by the system of political and social institutions under which they live*” (ibid. italics added). Of course, this should remind us – at least as an instructive contrast – of the central theme of the previous chapter: Kant’s belief (as thoroughly emphasized by Rawls, despite the objections of the “anthropological” Kantians) that one is capable of genuine moral freedom by virtue of one’s humanity and the concomitant possession of conscience, and *not* by virtue of one’s (arbitrary) political citizenship. Hegel, by contrast, holds the latter view, and does so directly in response to – *as a critique of* – Kant’s transcendentalism: freedom is realized not in the pangs of the abstract and isolated conscience but *in the world*, through political and social institutions at a particular historical moment (Avineri 1972, p. 65 and Taylor 1979, p. 51). Says Hegel in the *Preface to Philosophy of Right* in this vein: “The truth about Right, Ethics, and the state is as old as its public recognition and formulation in the law of the land, in the morality of everyday life, and in religion […] What it needs is to be comprehended so that the content which is already rational may also gain a rational form and thereby appear justified to free thinking” (Hegel 1991, Preface p. 3). Here, we
have an early articulation of the dual role of Hegelian political philosophy: first, to recognize that the most rational, freedom-realizing set of institutions already exists in the world – what is is rational – and, second, to articulate this fact in a way that is accessible to ordinary common sense; that is, to give philosophical knowledge a “rational form [...] justified to free thinking.”

Hegelian philosophy, like Rawlsian philosophy, refuses to judge the state against some distant ideal; instead, it is committed to telling us what the state is (Hegel 1991, Preface; see also Hegel 1964, p. 145, Avineri 1972, Preface, p. 4 and ch. 3 and Schwarzenbach 1991, p. 544). “Philosophy as reconciliation” is therefore the central theme of Rawls’s Lectures on Hegel.

But what, exactly, does “philosophy as reconciliation” mean? Here, Rawls’s Lectures draws an important contrast between reconciliation and resignation. The latter stance regards the prevailing institutional milieu as the least bad alternative: the existing social world may have a number of undesirable features, but is ultimately preferable to (more “unhappy”) alternative arrangements (Rawls 2000, p. 331). Reconciliation, by contrast, expresses a deeper commitment to our social world, which upon reflection is seen to express and protect our freedom – our status as ends. We have freedom, in other words, because political institutions recognize (and concomitantly institutionalize) “our dignity as persons who are free” (Rawls 2000, p. 331; see also Avineri 1971, p. 132). Consider, again, the instructive contrast to Kant, for whom adherence to the moral law requires that we raise ourselves above the contingencies of human nature (i.e. our particular desires and inclinations). In this way, we are all equally capable of freedom (as conscience-bearers). Hegel, by contrast, argues that the essential precondition for freedom is membership in a rational social world: genuine autonomy is dependent on (ultimately arbitrary) historical considerations (see the epigraph of the present chapter).

---

25 Hegel does not consider a rational social world a perfect social world. See, for example, Hegel 1991, §14 and §258A. See also Avineri 1972, p. 39: “Hegel’s argument is that only the acceptance of actuality as necessary, given the existing circumstances, can become a starting point for a new departure. This method is not an acquiescence in what is, but a critical understanding of it with a view to its transformation.”

26 Of course, this raises an obvious question: Hegel himself does not regard history as arbitrary. For him, as for Kant, nature is purposive – it tends towards rational ends, and so the purpose of philosophy is to vindicate Nature and history as ultimately rational. This is the metaphysical dimension of Hegel’s thought: the notion of Geist expresses Hegel’s belief that the course of history is fully intelligible and thus rightly subject to rational analysis. We will, for the most part, leave this metaphysical baggage aside: the following analysis is based on the assumption that Hegel’s political philosophy can be adequately understood without delving deeply into this metaphysical background. We will not (for the most part) be dealing with the evolution and emergence of Geist here. For this, see Hegel’s Science of Logic. Rawls also makes this interpretive assumption, and I follow him in doing so (Rawls
Why must philosophy popularize this fact? According to Hegel, the ethical life of the state is not real or actual until it is made self-conscious (Hegel 1991, Preface, p. 12 and §265A; see also Hegel 1977, §592, Taylor 1979, p. 25 and Rawls 2000, p. 332). The prevailing state of affairs, however, finds citizens alienated from this essential truth: “They tend not to understand,” explains Rawls, “that the social world before their eyes is a home” (Rawls 2000, p. 336; see also Hegel 1977, §582). This is the practical role of political philosophy (as reconciliation) – to make the ethical life of the state fully actual through reflection, judgment and, most importantly, through the development, and proliferation, of philosophical knowledge. It is only once the ethical life of the state is fully actualized (via reflection and judgment) that citizens come to adopt the preservation of their institutional milieu – for Hegel, this is another way of describing the common good – as their own end. This is connected, of course, to another central theme of Philosophy of Right: the notion that citizens are educated by (“liberal”) institutions to a conception of themselves as free – as free beings capable of willing an institutional system that respects their essential dignity (as free wills). It is to this important theme – to Bildung – that we now turn.

III.

What, exactly, does the “educative function” of political institutions mean? How do political institutions contribute to moral learning, as both Hegel and Rawls insist that they do? We begin, here, with Hegel’s Philosophy of Right27, which begins with an examination of the impact of

---

2000, p. 330). See also Schwarzenbach 1991, p. 541: “A Theory of Justice retains much of the fundamental structure of Hegel’s political theory while detaching this structure from its background metaphysics of absolute idealism […] Rawls retains the import of many of the most significant strands of Hegel’s metaphysics; Rawls does this, however, in what he takes now to be a practical, and no longer metaphysical, form.”

27 Unlike his interpretations of Kant and Rousseau, Rawls’s interpretation of Hegel’s moral and political philosophy is decidedly “mainstream”: even though Rawls treats Hegel’s metaphysics as separable from his politics – a potentially controversial assumption for some commentators – the Hegel described in Rawls’s Lectures is the straightforwardly liberal (and therefore familiar) Hegel, i.e. the Hegel who regards the fundamental purpose of political institutions as the protection of individual freedom and the overcoming of the atomism characteristic of market society. It therefore does not strike me as necessary or productive to scrutinize the idiosyncrasies of Rawls’s Hegel: our primary guide in the sections that follow is the Lectures on the History of Moral Philosophy, which is
principle-guided institutions on the moral development of citizens. In that work, Hegel describes the actualized – or, genuine – form of freedom as the unity of form and content (Hegel 1991, §9). Form here corresponds to the first “moment” of abstract universality – the generalized capacity for agency – while content corresponds to the second “moment” – the particularized expression of the capacity for choice. Form, then, is a state of pure, unconditioned indeterminacy: Hegel asks us to imagine the content-less consciousness (ibid, §5; see also Wood 1990, p. 398). “I can make myself,” says Hegel in an 1824 lecture, “entirely empty” (quoted in Rawls 2000, p. 337). Indeed, an important expression of the freedom of the will is its ability to detach itself from every particular aim, impulse or inclination; free will is conceived at this stage as the purely negative capacity to reject.

Upon willing, though, consciousness evolves from indeterminacy to determinacy – consciousness gives itself that missing content by positing an end worth pursuing; the adoption of, and identification with, ends is the second “moment” along the path to actualized freedom. What, then, is the appropriate content for the free will to will? For Hegel, the content of the will expresses itself in one of two ways. On the one hand, the individual arbitrarily chooses from its essentially random instincts, desires and inclinations. This inadequate form of freedom is characterized by the natural will’s lack of inner organization based on articulated principles. Says Hegel: “An impulse is simply a uni-directional urge and thus has no measuring-rod in itself, and so this determination of its subordination or sacrifice is the contingent decision of the arbitrary will” (ibid., §18). The natural will does not possess the means of giving priority to certain ends over others; hence, the arbitrary character of its content (ibid., §11). A free will, in other words, must will something appropriate for a free will.

Such a will is only implicitly, or formally, free: despite exhibiting the capacity for choice – that is, the capacity to accept or refuse an impulse – the formally free will’s dependence on impulse occasionally supplemented with the work of Avineri, Schwarzenbach and Taylor, all three of whom are important interpreters of Hegel and all of whom fundamentally agree with the liberal portrayal of Hegel’s political philosophy. Ultimately, our purpose in the remaining parts of this chapter is to get a better understanding of Hegel’s Philosophy of Right and to expose the extent to which A Theory of Justice and Political Liberalism are deeply indebted to that earlier text – indebted to a degree unappreciated by the vast majority of Rawls’s commentators and critics.
and inclination for its choice contradicts its possible self-determination. Such a will must choose from something *given* (ibid., §15). Hegel is critical of Kant’s response to this problem: Kant regards this arbitrariness of will as a fixed condition and so, to attain moral freedom, the will must abstract from all of its particular inclinations (ibid., §15). Freedom, for Kant, consists in willing only universal content. For Hegel, on the other hand, the natural will is *not* a fixed condition: it is transient, and the inclinations one finds in oneself are not necessarily opposed to freedom; they *can* be made rational – systematized. There is, in other words, a rationality immanent in the immediate inclinations (ibid., §19). They must, however, be “purified;” individuals must free themselves from the “immediacy” of the inclinations (ibid., §20).

The means to this transcendence – or, “purification” – is the conscientious organization of our natural inclinations and ends according to rational principles: “Reflection invests [the impulses] with abstract universality and in this manner purifies it from its crudity and barbarity” (Hegel 1991, §20). This is the moment where the inclinations lose their form as simply found; instead, conscientious reflection is what makes the choosing of this or that inclination fully rational. Freedom, in other words, is the principled (or systematic) organization of the impulses (ibid., §21). And this is precisely the connection between Hegel’s moral and political theory: the content, or “purposes,” of the individual will flow from a coherent political system based on the principle of mutual recognition (ibid., §150). “The system of right,” in other words, “is the realm of freedom made actual” (ibid., §4 italics added). Says Rawls of the uniquely Hegelian confluence of the moral and political: “Hegel wants us to find our moral compass in the institutions and customs of our social world itself, as these institutions and customs have been made a part of us as we grow up into them and develop habits of thought and action accordingly” (Rawls 2000, p. 333; see also Hegel 1991, §150). “The rational destiny of human beings,” says Hegel in this vein, “is to live with a state” (Hegel 1991, §75 italics added).

The particular, or natural, will is thus brought back to universality in light of its guidance by the principle of mutual recognition. And our awareness of this principle – as well as the desire to have our actions conform to its dictates – is dependent on its being embedded in prevailing political and legal institutions; this is what we mean when we speak of the educative function of political institutions and of moral learning taking place in a political-institutional context. After all, the movement from universality to particularity to mutual recognition depends on the prior existence of the proper institutional background (Hegel 1991, §4; see also Rawls 2000, p. 339).
Here, we must again stress the difference from Kant, for whom the moral law is transcendental and can never be fully institutionalized in a political milieu. Kant’s kingdom of ends is a regulative ideal that can never be achieved in practice, whereas Hegel’s ideal state is already guided by the (rational) principle of mutual recognition (as institutionalized via political activity).28 This is precisely the meaning of the first half of Hegel’s famous (and famously misunderstood) maxim: “What is rational is actual.” Says Rawls of this important point: “For Hegel, in contrast to Kant, the aim of the account of ethics is not to tell us what we ought to do – we know that – but to reconcile us to our real social world and to convince us not to dwell on our real world’s shortcomings” (Rawls 2000, p. 334).

On Rawls’s view, Hegel’s fundamental problem with Kantian ethics is that he views Kant as moved by a “desire for radical purity” – by the desire to act from the moral law and nothing else (ibid., p. 335). This way of viewing ourselves is dismissed by Hegel as both “narrow” and “alienating;” narrow because it forces us to disregard essential aspects of ourselves as members of a particular socio-political community and alienating because it excludes “so many of the desires and aspirations of everyday life […] [Hegel] wants to allow that the aims of love and friendship, family and association, are fully consistent with ethical life” (Rawls 2000, p. 334; see also Hegel 1991, §33). Here, we return to the pedagogical theme with which we began this section: the whole ensemble of rational (i.e. freedom-providing) social and political institutions – this is the location of ethical life – teaches us not only our station in life, but the concomitant duties implied by that role. Says Hegel: “In an ethical community, it is easy to say what man must do, what are the duties he has to fulfil in order to be virtuous: he has simply to follow the well-known and explicit rules of his own situation” (Hegel 1991, §150).

---

28 This important difference is related to the contrasting – no, irreconcilable – epistemologies of Kant and Hegel. Indeed, part of Hegel’s purpose in writing Philosophy of Right (among other works, including the Logic) is to transcend the limits that Kant imposes on reason (Hegel 1991, §26A). In contrast to Kant for whom the noumenal thing-in-itself is permanently unknowable – if all that we “know” about the natural world is the formed by our thought processes, it follows for Kant that we can never have immediate experience of reality – Hegel believes that the world is fully knowable because it is the product of human reason historically realized in the phenomenal world; hence the title of Hegel’s first attempt to systematize his philosophical views: The Phenomenology of Spirit. See, for example, Guyer 1993, p. 171 and p. 189, Avineri 1972, p. 64 – 65, Taylor 1979, p. 47 – 53 and 124 – 125 and Kauffman 1988, ch. 3.
Duty, in other words, does not require the transcendence of the world or of the inclinations; one can fulfil one’s duty, on Hegel’s view, by acting in accordance with the rules, practices and expectations of one’s concrete situation. Here, Hegel devalues the importance of the abstract capacity for agency – this is where Kant’s emphasis lies – and instead advocates respect for the determinate individuality of persons. Says Hegel: “Duty […] is liberation from indeterminate subjectivity which, never reaching reality or the objective determinacy of action, remains self-enclosed and devoid of actuality” (ibid., §149). Adds Rawls: “As individuals we know our place in the social world; we are fulfilled, and gain our happiness, by doing our duties well […] there is no essential conflict between duty and happiness or duty and inclination. Part of Kant’s error is to think that there might be” (Rawls 2000, p. 352 italics added; see also Taylor 1979, p. 20 and p. 80). Moral duties are thus given by the lived (and evolving) forms of ethical life – the rules, values, and practices, as well as the concomitant obligations, of our socio-political association (Sittlichkeit). Here, Hegel reverses the order supposed by Kant, for whom moral duties ought to constitute community. But this leaves us with an obvious question: How does Hegel reconcile freedom with adherence to rules and practices that make up ethical life?

To answer this important question, we return to an earlier question: what is the appropriate content for a free will to will? According to Hegel, the free will must adopt itself – its freedom – as its proper object and end (Hegel 1977, §598). And this form of self-determination has a specifically political dimension (Hegel 1991, §10). Indeed, the free will wills itself as an actualized free will – it is autonomous – when it wills a system of social and political institutions based on the principle of mutual recognition – a community, in other words, in which the will – all wills – can will and act freely (ibid., §111). This is the meaning of the following passage from Philosophy of Right: “The absolute determination […] of the free spirit is to make its freedom into its object […] The […] Idea of the will is the free will which wills the free will” (Hegel 1991, §27).

Again, Hegel: “Right is something utterly sacred, for the simple reason that it is the existence of the absolute concept, of self-conscious freedom” (ibid., §30).

---

29 The implication of this view of Sittlichkeit, according to Hegel, is that modern society no longer has need for the virtue (or culture) of heroism. Instead, virtue is understood as rectitude (Hegel 1991, §150). See also TJ, §19.
The system of right, in other words, is the *potential* freedom of the will *actualized*. And so, when we abide by the moral duties implicit in the ethical life of our community – when we abide by *right* – it is precisely an expression of our freedom; after all, the rational community is the appropriate object – the *only* appropriate object – for a free will to will (Taylor 1979, p. 81). For Hegel, then, the rational state makes possible the free play of personality – via the protection of individual rights – and is the essential expression of free personality (ibid., §28 and §147). Duty, in other words, remains a form of *self*-legislation (as it was for Kant, although in an unsatisfactory form).

Free personality is thus reconcilable with devotion to the common good: individuals recognize the laws as the essential source of their freedom, and so are committed to their preservation, just as they are committed to their own particular reflective goods. We appropriately adopt the preservation of our institutions as an end, just as we adopt the preservation of our own freedom as an end; indeed, these imperatives are inextricable. In Hegel’s language, this is another way of describing the mutual recognition structure of the particular and universal wills. Ethical life, in other words, expresses the necessary interdependence of the universal and particular goods. In Rawls’s more familiar language, the good is defined as the actualization (and preservation) of right through the conscientious, self-determined activity of individuals (ibid., §136-138; see also Schwarzenbach 1991, p. 552).

Indeed, Hegel believes that individuals realize their essential nature by acting as a vehicle for the realization of the universal will; the individual gains worth, in other words, through action for the sake of the political order (ibid., §140). Here, Hegel *seems* to advocate the primacy of the universal over the particular, and thereby risks upsetting the delicate balance that is supposed to characterize the mutual recognition structure of spirit. And surely no liberal can accept the view that *political* activity is the quintessential vision of the good life (Rawls 2000, p. 369). But Hegel does not endorse such a view: he emphasizes the Greek polis’s negation of individual particularity as the reason for its demise, and associates the (necessary) emergence of self-consciousness with the individual assertion of ends-status (Hegel 1991, §148, §150 and §261A; see also Hegel 1964, p. 163-164, Hegel 1971, p. 154-155, Avineri 1972, ch. 2 and Taylor 1979, p. 8 and p. 97). Of course, this repudiation of community is a necessary stage in the emergence of ethical life: eventually, the rational state will strike a tenable balance between atomistic self-
assertion and all-encompassing patriotism. As we shall see in detail below, this clearly resonates with Rawls’s own political project.

In the end, Hegel’s ideal state is characterized by a common trust in, and devotion to, the objective political order (ibid., §2). “Citizens,” adds Rawls, “knowingly and willingly acknowledge this universal (collective) interest as their own, and give it the highest priority” (Rawls 2000, p. 355). Indeed, without validation from citizens, the law remains characterized by its one-sided universality; the legal and political institutions of the state must receive recognition from independent selves, who are in turn recognized by the law of the state. Says Hegel in this vein: “The state is actual only when its members have a feeling of their own self-hood […] If their subjective aims are not satisfied, if they do not find that the state as such is the means to their satisfaction, then the footing of the state is insecure” (Hegel 1991, §265A italics added; see also §268A). This is the moment when the ethical life of the state becomes fully actualized – when citizens self-consciously recognize the freedom-guaranteeing character of their institutional milieu and, in turn, act in the service of its preservation.

That the state can achieve this level of actuality attests to the hospitality of the natural world to the actualization of freedom. Reason, for Hegel, is immanent in the world, in relations of contract, property, marriage, family, and citizenship (Hegel 1991, §77). The rights of property, for example, show respect for (moral) personhood and for freedom by institutionalizing our ability to appropriate, develop and alienate things, over which persons rightly have control (ibid., §39 - §45, including the Remark).³⁰ To be denied property rights, in other words, is to be denied the rights of personality; indeed, Hegel rejects any justifications of property rights based on utilitarian or welfarist assumptions (ibid., §37 and §53; see also Rawls 2000, p. 343). Property rights are explicitly conceived by Hegel as a vehicle for the preservation of particularity and of self-realization; it is only once he develops a more robust conception of ethical life that Hegel endorses the limitation of the right to unlimited accumulation, which he initially accepts in the

---

³⁰ Here we have an example of Hegelian philosophy in action, exposing the hidden inner rationale behind seemingly commonplace institutional patterns.
name of personality rights, but eventually rejects as contrary to the public good (ibid. §49 and §107). 31

All of this captures the meaning of the second half of Hegel’s famous (and famously misunderstood) maxim: “What is actual is rational.” Of course, not everything for Hegel is rational: only those institutions that manifest the principle of mutual recognition – institutions, in other words, which are reasonable – achieve actuality (ibid., §1). The important difference between “existence” and “actuality” has been a fundamental cause of confusion among Hegel’s interpreters. Indeed, as we have already seen, Hegel does not equate a rational social world with a perfect social world (ibid., §258). 32 And Rawls goes to great lengths to emphasize the important – and, according to Hegel himself, the insuperable – distance between the two: to the problem of “der Pöbel” (the poverty-stricken rabble inevitably created by the market society), for instance, there is no satisfying solution (ibid., §185, §236 and §245; see also Rawls 2000, p. 348, Avineri 1972, p. 147 – 154 and Wood 1990, p. 247 – 255).

According to Rawls, then, Hegel is a “moderately progressive reform-minded liberal, and […] his liberalism is an important exemplar in the history of moral and political philosophy of the liberalism of freedom” (Rawls 2000, p. 349). What, exactly, is a “liberalism of freedom” and in what ways does Rawls’s own liberalism take its cue from Hegel? By “liberalism of freedom,” Rawls means that the “first principles” of Hegel’s political philosophy “are principles of political and civic freedoms” (Rawls 2000, p. 330). As we have already seen, freedom here is an institutional concept: Hegel’s “liberalism of freedom” is ultimately characterized by its emphasis on a system of social and political institutions that make possible, and protect, citizens’ basic freedoms (ibid., p. 352). Says Rawls: “A Theory of Justice follows Hegel when it takes the basic structure of society as the first subject of justice” (Rawls 2000, p. 366; see also PL §VII).

31 See also Hegel 1991, §112-113 and §127-128. In these sections, Hegel speaks of the emergence of positive duties of concern for others (as oppose to negative duties of non-interference). Indeed, morality requires the pursuit not just of our own freedom, but the freedom of all (see below). This is also connected to Hegel’s discussion of welfare rights in §114 and §129-§130: we have a duty, which is correlative to the rights of personality (or self-determination) of our fellows, to secure the material conditions necessary for genuine (or actual) freedom. This is the only legitimate constraint on property accumulation.

32 Indeed, there is also a critical (anti-status quo) dimension to Hegel’s account of philosophy: the task of philosopher, in part, is to convert existence into actuality (Hegel 1991, Preface). The Philosophy of Right itself acts as a bridge between the de facto, historical state and the actualized, rational state.
IV.

Just as there are three “moments” in the process of the actualization of the free will – indeterminacy to determinacy to the principled organization of the will’s determinate content – so too are there three “moments” in the emergence of *Sittlichkeit* or ethical life: family, civil society and the state. According to Hegel, then, the family is ethical life in its natural or immediate form; it is based on love and feeling, not reason and calculation (Hegel 1991, §176). Family members do not have any status as independent selves – the family, for Hegel, represents the wholesale rejection of atomism – and individual worth is determined by the outcome of the family union (ibid., §158). The family, in other words, is the first foundation of community: it is understood by Hegel as a kind of collective “person” insofar as the individual wills of its members are sublimated in the common project (ibid., §163 and §170). The family is also an important locus of education: children are rightly prepared by their parents for citizenship and for participation in the wider affairs of the political community (ibid., §174). They are educated to an idea of themselves as self-determining ends (ibid., §175). Eventually, though, the substantial unity of the family gives way to the presence of multiple families and, in turn, to the abstract or formal unity of civil society (ibid., §181). For Hegel, civil society represents the assertion of the individual outside community; and it is only with the introduction of the constitutional state committed to the protection of rights that we can return to a modified form of the unity that characterized the family.

According to Hegel, then, civil society is comprised, among other things (such as the administration of justice and the police), of the market or the “system of needs” (Hegel 1991, §189). This is a familiar image: individuals exchange goods in order to fulfil their wants and needs, and these are constantly evolving as the economy itself develops (ibid., §187, §190, §191A and §195; Avineri 1972, p. 92 and 144 – 145). Eventually, classes are formed and the division of labour becomes more sophisticated (Hegel 1964, p. 190). For Hegel, participation in civil society represents the conscientious separation of persons into public and private selves: we no longer see ourselves as members of a community, but as individuals with private interests in need of fulfilment. *This* is the characteristic mark of the *modern* state, which is now conceived as a (mere) instrument for the protection and defence of property and contract relations (ibid., p.
Civil society is the sphere of universal *egoism*. But for Hegel this system is *not* exclusively defined by this fulfilment of needs and wants and by the competitive individualism that generally characterizes market society. Just as property rights have an ethical function – recall the connection Hegel draws, discussed above, between *property* rights and the rights of *personality* – so too does the market (Hegel 1991, §184). Eventually, that is, individuals come to recognize (after time and reflection) their *mutual interdependence*; we are not isolated atoms, says Hegel, but are in need of the goods and services provided by our fellows (just as they are equally in need of our goods and labour) (ibid., §199; see also Rawls 2000, p. 346 and Avineri 1972, p. 90 – 91 and p. 146 – 147).

In this way, the market functions as an *essential precondition* of the rational state (ibid., §189 and §238A). Says Hegel in this vein: “The principle of modern states has enormous strength and depth because it allows the principle of subjectivity to attain fulfilment in the *self-sufficient extreme* of personal particularity, while at the same time *bringing it back to substantial unity* and so preserving this unity in the principle of subjectivity itself” (ibid., §260). In other words, market relations, which are initially characterized by the pursuit of narrowly conceived self-interest, eventually lead market actors back to the aims of the state by forcing us to reconsider our relations to one another – our fundamental interdependence. *This* is the market’s ethical function: it draws us back towards our fellows while simultaneously giving each citizen the freedom to fulfil his or her self-interest. It strikes the balance characteristic of ethical life – between the particular and the universal.

But it is only via the “*Constitution of the State*” that individuals achieve the substantial (or actualized) unity that is the aim of Hegel’s political philosophy (ibid., §157). Whereas the unity

---

33 See also this important passage from the same section: “The universal does not attain validity or fulfilment without the interest, the knowledge and the volition of the particular […] individuals do not live as private persons merely for these particular interests without at the same time directing their will to a universal end and acting in conscious awareness of this end” (Hegel 1991, §260).

34 It must be noted, here, that for Hegel this transition (from self-seeking market behaviour to conscious interdependence) was hardly seamless; indeed, Hegel clearly recognizes the potentially stultifying effects of the modern division of labour. An important part of his project – the role he assigns to the state, as well as the reason for his advocacy of the corporation (see below) – is to overcome the (class) polarization of society that undermines any genuine sense of community. For instance, the state must regulate economic activity in order to mitigate extreme inequality (see the discussion of property rights above). On the problem of alienated labour (and the concomitant class conflict) in Hegel, see Avineri 1972, p. 87 – 109 and Avineri 1970a, ch.1.
of the family is based on love and the unity of civil society remains, for the most part, unconscious, it is only in the state that reason becomes fully conscious of itself (ibid., 257). This view is most clearly stated in the following passage from Philosophy of Right, worth quoting at length: “The state is the actuality of concrete freedom. But concrete freedom requires that personal individuality and its particular interests should reach their full development and gain recognition of their right and also that they should, on the one hand, pass over of their own accord into the interest of the universal, and on the other, knowing and willingly acknowledge this universal interest even as their own substantial spirit and actively pursue it as their ultimate end” (ibid., §260 italics added). There is a kind of triple movement at work here: first, Hegel emphasizes the state’s necessary commitment to the protection of individual rights and freedoms; the effect of this constitutional commitment, second, is the eventual attachment of the individual beneficiaries of state activity to the preservation and vitality of their shared institutional milieu. 35 And then, third, this common good – understood, here, as shared political activity in the service of institutional stability – acquires a kind of independent life. Bourgeois become citizens. 36 This transformation is the fundamental precondition for the emergence of ethical life and is itself ethical life. Avineri describes this phenomenon as the “conscious identity of the subject and the state […] [it is] a condition of the adequate functioning of the commonwealth” (Avineri 1972, p. 178).

Says Hegel in this vein: “Individuals do not live as private persons merely for these particular interests without at the same time directing their will to a universal end and acting in conscious awareness of this end” (ibid.; see also Avineri 1972, p. 7). Again, it is not the state’s provision of the freedom necessary for the pursuit of self-interest that arouses citizen interest; the community whose survival depends on the satisfaction of narrow wants and needs is of no interest here nor is its survival likely. Instead, the stable community is animated by what Rawls

35 Using the language of A Theory of Justice, we can say something here about the inextricability of the “benefits and burdens” of social cooperation (TJ, §19). This is the subject of a lengthy discussion below. Indeed, the remainder of this chapter is committed to examining the impact of this Hegelian triple movement on Rawls’s own thought.

36 This same movement plays itself out on the political plane with the state’s equal recognition (and integration) of estates and corporations: these units of political organization are based on a particular commonality (of class consciousness) among members (whether labourers, middle class or bureaucrats), and are sustained by the equal recognition (and the concomitant participation) of state institutions. See Hegel 1991, §203 and Avineri 1972, ch. 8.
calls a “common-good conception of justice” (Rawls 2000, p. 355). There is for Hegel something independently – *intrinsically* – meaningful about the universal labour required for the preservation, evolution and flourishing of public institutions. Says Rawls of this crucial difference: “The state makes one a *citoyen* and not simply a *Bürger* […] What raises human life above the workaday *bügerliche* is the recognition of the universal interest of all citizens in participating in and maintaining the whole system of political and social institutions of the modern state that make freedom possible. Citizens knowingly and willingly acknowledge this universal (collective) interest as their own, and they give it the highest priority. They are ready to act for it as their ultimate end” (Rawls 2000, 355). Adds Avineri: “What Hegel is looking for is […] the consciousness of belonging to a community, that feeling which would not view the community in instrumental terms” (Avineri 1972, p. 84). Finally, Taylor: “We have gone beyond the opposition of self-goal and other-goal […] the individual defines [her] identity principally by the public experience of the society” (Taylor 1979, p. 86).

V.

Having thus outlined the themes and concepts characteristic of Hegel’s moral and political philosophy, we turn now to the Hegelian dimensions of justice as fairness. The essential theme of Rawls’s *Lectures* on Hegel is the “deep social rootedness of people within an established framework of their political and social institutions” (Rawls 2000, p. 366). As we have already seen, Hegel regards this sense of embeddedness as a necessary corrective to the abstract, radically purist orientation of Kant’s ethics. It is the purpose of the remainder of this chapter (and of the dissertation in general) to examine this theme – the social dimension of human nature and of the moral sentiments – in Rawls’s own political philosophy. This marks an important turning point: the image of Rawls’s political philosophy as individualistic and atomistic is no longer tenable, and will be exposed as a serious, and yet enduring, misinterpretation of Rawls’s thought (see e.g. MacIntyre 1981, Sandel 1982, Buchanan 1989, MacIntyre 1989, Taylor 1989, Haldane 1996, Miller 1999, Simhony and Weinstein 2001, Mullhall and Swift 2003 and Taylor 2003). As we shall see in this and the next chapter, Rawls’s political philosophy is deeply connected to some fundamentally republican themes, including the duty of political participation,
the importance of civic virtue and, most importantly, an emphasis on properly designed political institutions as the source of freedom.

In *A Theory of Justice*, for example, Rawls speaks of the duty of justice – the duty, that is, to comply with and support just institutional arrangements, and to contribute to the emergence and evolution of possible future arrangements – as a “fundamental natural duty” irrespective of one’s voluntary acts (i.e. having “signed” some kind of original contract) (*TJ*, §19). “We are not to gain from the cooperative efforts of others,” says Rawls, “without doing our fair share” (ibid., §52). What we shall ultimately find is that a prominent though unappreciated theme of *A Theory of Justice* – and of Rawls’s entire oeuvre – is determining “how we become bound to one another” (ibid., §51 italics added). Let us begin this inquiry by returning to the theme of education and moral development, which are discussed by Rawls in Chapter VIII of *A Theory of Justice*. In that chapter, Rawls (like Hegel) emphasizes the educative role of social and political institutions in the acquisition of proper moral sentiments. Indeed, Rawls appears to be outlining a theory conspicuously reminiscent of ethical life: his account begins with the family, emphasizing its character formation function; he then turns to our (unreflective) participation as individuals in civic associations and, finally, to the process whereby we become attached to the principles that inform our institutional milieu. There are, in other words, three essential “moments” (or, as Rawls calls them, “laws” or “tendencies”) in the development of the Rawlsian sense of justice. The Hegelian echoes of Part III of *Theory* are undeniable.

Like Hegel, for whom the first moment of ethical life takes place in the family, Rawls begins his discussion of moral development in early childhood: “I assume that the sense of justice is acquired gradually by the younger members of society as they grow up” (*TJ*, §70 italics added). In fact, Rawls regards the family as part of the basic structure of society – just as the principles embedded in our institutional milieu have a pedagogical function, so too do the rules, practices and values that characterize family life. The complication, though, is that the child’s instincts are primitive. Unlike the mature citizen who can intuit (or question) the rational content of the laws, the child is unable to question the rules and injunctions of her parents, the authority figures in her
life.\textsuperscript{37} The concept of justification is foreign at this stage, and so she must follow them without comprehending the overarching (moral) system they are attempting to teach her; she feels guilt when, inevitably, she transgresses a particular rule and is rebuked accordingly. Such guilt is the unconscious manifestation of the child’s recognition of her parents as object worthy of admiration and emulation.

For the same reason – namely, underdeveloped reason – any psychological development must necessarily be the result of affective rather than rational attachments. The emergence of affectionate feelings towards one’s parents is not the result of any rational, instrumental calculation – egoistic self-interest, for instance – but rather a recognition, and eventually a reflection, of the love and attention the child is already receiving from her parents: “The child comes to love the parents only if they manifestly first love [her]. Although the child has the potentiality for love, [her] love of the parents is a new desire brought about by [her] recognizing their evident love for [her] and [her] benefiting from the actions in which their love is expressed” (ibid.).\textsuperscript{38} These benefits include unconditional love, a sense of security and, most importantly, the affirmation of her sense of self-worth. Indeed, the child’s budding sense of self-esteem – the thought of herself as an end – is the direct result of the affection and encouragement of her parents (ibid.). This marks an important moral milestone in the life of this future citizen: namely, the capacity to form lasting, meaningful attachments based on sentiments of fellow feeling and even love (\textit{TJ}, §70; cf. Hegel 1991, §175).

Whereas this first stage of psychological development is generally based on unrecognized moral precepts, the second stage, the “morality of association,” is instead based on public standards and

\textsuperscript{37} For the sake of simplicity, Rawls assumes that said rules and injunctions are governed by considerations of both reason and justice, though he does acknowledge that this is of course not always the case. See Rawls, \textit{Collected Papers}, p. 101. This also recalls a prominent debate in Rawls scholarship: Rawls’s professed non-application of the two principles of justice to the family, and the concomitant disadvantage this poses to its female members. See Okin 1991, Ch. 4. Despite this lacuna, the educative dimension of family life is central to Rawls’s argument about the emergence of the sense of justice.

\textsuperscript{38} Rawls credits Rousseau’s \textit{Émile} with the initial articulation of this psychological law: “A child is naturally disposed to kindly feelings because he sees that everyone about him is inclined to help him, and from this experience he gets the habit of a kindly feeling towards his species” (\textit{Émile} 174). See also Rawls, “The Sense of Justice” in \textit{Collected Papers}, p. 96. Of course, Rawls fails to recognize the potential converse tendency: namely, the painful psychological experience of loving one’s parents without feeling loved in return. In what follows, I leave this possibility aside.
expectations (TJ, §71). Rawls characterizes these as the “common sense” rules (and expectations) that a particular person in a particular position must follow and fulfil in order to ensure the continued existence of the association in question. Examples of these kinds of associations include schools, neighbourhoods, business associations, unions and, eventually, the national community (ibid.; cf. Hegel 1991, §150). Just as one has learned the virtues of a good son or daughter – in light of the expression of parental approval or disapprobation as well as different, slowly learned conceptions of the ideal child – one is now ready to learn the virtues of a good student, neighbour, team member, etc. It is precisely this process of recognizing and fulfilling the ideal conception of each particular role – including the aims and purposes of the role, as well as of the association of which it is a part – that contributes to the process of moral learning: for Rawls, this requires “increasingly greater intellectual judgment and finer moral distinctions” (ibid.). To use Hegelian language, what Rawls is describing is the constitution and eventual recognition of duty as determined by the lived forms of ethical life, and our determinate position within it.

Eventually, as we move through different roles within one or many associations, we come to recognize the overarching system governing our association and to an appreciation of each member working towards its continued realization – we recognize our fundamental interdependence. “In due course,” says Rawls, “a person works out a conception of the whole system of cooperation that defines the association and the ends which it serves. He knows that others have different things to do depending upon their place in the cooperative scheme” (ibid.; cf. Hegel 1991, §260A). This second stage is thus defined by the eventual development of our capacity for sympathy; in fact, the extent to which we are able to regard things from other people’s perspective is a quintessential expression of our moral sensibility.

As in the first stage of psychological development, Rawls places an important caveat upon the inner workings of this second stage. The public rules of the association in question must be known by all to be (reasonably) just: they must respect the freedom and equality of all its members, all of whom must benefit, and know that others benefit, in accordance with their roles; all must come to a mutual recognition of their duties and perform them as an act of good will. Eventually, the satisfaction of the previous two conditions will arouse sentiments of fellow feeling and mutual trust (Rawls 1999d, p. 102 – 105). In fact, only in such cases will the development of our moral attitudes and sensibilities translate into effective and stable political
associations: “In due course the reciprocal effects of everyone’s doing his share strengthen one another until a kind of equilibrium is reached […] When the moral ideals belonging to the various roles of a just association are lived up to with evident intention by attractive and admirable persons, these ideals are likely to be adopted by those who witness their realization.” (TJ, §71). Just as the child develops affectionate feelings towards her parents, so too does a bond of friendship and mutual trust develop among associates sharing just institutions. Of course, there is one essential difference: whereas the child’s bond is affective, the bond developed in the second stage of moral and psychological development is instead based on the rational recognition of mutual benefit and a common conception of justice. It is thus that the moral capacity for “justice and fairness, fidelity and trust, integrity and impartiality” continues to develop (ibid.).

The culmination of this process – the third stage, which Rawls dubs “the morality of principles” – is defined by the complete emergence of a sense of justice. Here, one can witness a fundamental moral transformation: from an appreciation of a particular association as well as the benefits one receives from active participation in it towards an allegiance to the first principles governing the association. This attachment to shared justice principles is the essential source of stability in any political community. Says Rawls: “[T]he recognition that we and those for whom we care are the beneficiaries of an established and enduring just institution tends to engender in us the corresponding sense of justice. We develop a desire to apply and to act upon the principles of justice once we realize how social arrangements answering to them have promoted our good and that of those with whom we are affiliated. In due course we come to appreciate the ideal of just human cooperation” (TJ, §72). Rawls describes this process as the transition from being “a good sport” to being “a just person” (ibid.). Full maturity, in other

39 There are, according to Rawls, two tendencies that undermine the stability of large-scale, public goods-producing social arrangements: first, when (self-interested) individual members are tempted into shirking their public responsibilities, such as paying income tax (despite having benefited from the provision of public goods); and second, when the prevailing social mind-set is such that members with a normally effective sense of justice fear that others will likely not do their part. It is dangerous (or unfair or unattractive) to abide by the rules, in other words, when our fellows can’t be trusted to do the same (even though they may, in the end, do so). Of course, if a society can erase the first kind of temptation through public institutions – such as the permanent prospect of punitive sanctions – the second will disappear in turn. This, according to Rawls, is the Hobbesian solution to the problem of social cooperation (TJ, §52). With his emphasis on mutual trust and fellow feeling, Rawls (like Hegel) rejects this atomistic view despite the centrality of the Hobbesian social contract trope in Theory, “for when these attachments exist, [reasons other than self-interest, expediency or fear] are acknowledged for acting fairly” (ibid., §74).
words, requires the reorganization of one’s ends and inclinations in light of the principles of justice implicit in one’s prevailing social-political milieu; one can, by extension, make sense of one’s moral guilt in light of well-defined justice principles (as opposed to the vitality or deterioration of one’s association) (TJ, §73 and §74 and PL §I.1.4 and McClennen 1989, p. 22 – 24 and p. 30).

With the conclusion of this (early) account of stability in a well-ordered society, we have come to a central controversy in Rawls scholarship: the supposed shift from A Theory of Justice to Political Liberalism – for some, it is monumental (Ackerman 1994, Barry 1995 and Estlund 1996). Indeed, the impression one gets from Rawls’s Introduction to the latter work seems to call into question this chapter’s emphasis on these three “moments” in the moral development of the citizen (although, in an attempt to highlight Hegelian themes in Rawls’s normative work, such an endeavour remains worthwhile). I do not accept such a view. What I hope to show, in the end, is that the ideal of stability outlined in Political Liberalism is a “recasting” – as opposed to the wholesale abandonment, say – of the earlier account of A Theory of Justice, and is regarded by Rawls as an “internal” revision of justice as fairness (PL, Introduction, p. xiv – xvii and p. xxix).

The fundamental problem with Part III of Theory, according to Rawls, is this: the “morality of principles” – citizens’ dedication to the principles implicit in their liberal institutional milieu, which renders our community stable – fails to take into account the possibility of competing allegiances. Citizens, in other words, are treated as perfectly open to the principles of justice as fairness; to use the language of Political Liberalism, justice as fairness is assumed to be their comprehensive doctrine (ibid., p. xvi; see also Scheffler 1979, p. 399 – 401). The reality, however, is different: citizens may have other comprehensive doctrines – other religious, moral or metaphysical worldviews – that are fundamentally hostile to Rawls’s liberalism (or to any form of liberalism for that matter). Adopting these justice principles as the ground of one’s action may therefore be experienced as a kind of sacrifice; we might have to give up an essential part of ourselves – our most deeply held beliefs in the most important matters imaginable – in
order to participate in political affairs. We thus run the risk of alienating certain citizens from the wider (political) community. In other words, Rawls indicts his own *Theory* for overlooking the fact of pluralism.

According to the account presented below and in later chapters, there is no significant (irreconcilable) tension between the earlier *A Theory of Justice* and *Political Liberalism*. Indeed, *Political Liberalism* is best understood as an expansion upon *A Theory of Justice* – namely, the incorporation of a more sophisticated account of the psychology of the individual citizen – as opposed to a wholesale rejection. In this way, the “morality of principles” is still an operative consideration at the time of *Political Liberalism*, but it now faces a more difficult process of acceptance: it must be seen as fully reconcilable with one’s particular comprehensive doctrine. But, in the end, the content of the two principles remains the same, and so too does the importance of citizens’ commitment to those principles (*PL*, §VII; see also Hill, Jr. 1994, p. 333 – 336). Regardless of one’s comprehensive doctrine, Rawls’s two principles of justice reflect the (shared, not peculiarly Kantian) intuition that the purpose of political institutions is (at least in part) to assure favourable conditions for the realization of basic human potentialities (i.e. rational agency and the development of a sense of justice) (Cohen 1993, p. 278 – 281).

That is why Rawls’s discussion of political education in 1993, to which we shall turn shortly, sounds much the same as the discussion of the “morality of principles” in 1971; and so when, in the Introduction, Rawls calls the “political conception of the person” part of an “extensive revision” it’s not entirely clear what this means – it doesn’t strike me as much of a revision at all (and, in a footnote from *Political Liberalism*, Rawls specifically references that earlier account). Indeed, the “political” conception of the person is precisely the acceptance of Rawls’s justice *principles* in the manner described in *Theory*. The main difference, in the end, is

---

40 On the question of unjust or intolerant comprehensive doctrines, see PL, §V.5 and §V.6. Ultimately, Rawls sees such doctrines disappearing from the society of justice as fairness: “It is surely impossible for the basic structure of a just constitutional regime not to have important effects and influence as to which comprehensive doctrines endure and gain adherents over time […] We must accept the facts of commonsense political sociology.” We take up this question in detail in Chapter 5.

41 See also Rawls 2001, §55.1, where Rawls discusses the question of stability and the special psychologies (i.e. envy and spite, as discussed in Chapter 1). After referencing the aforementioned account of *Theory* – and of the morality of principles – Rawls offers the following: “I would not change [the account of *Theory*] substantially […] What is said [in the *Restatement*] supplements that account.” See also Weithman 1999, p. x – xi.
because citizens are now understood to hold “intermediate” comprehensive doctrines, the transition from the “morality of association” to the “morality of principles” is much less seamless that Rawls initially posits (in Theory). The comprehensive doctrine holder must, in a sense, be won over by Rawls’s own political conception of justice (Cohen 1993, p. 273). But this movement is no less essential to the stability (and vitality) of our political community. This is captured by Rawls’s emphasis on the reasonableness of the plurality of comprehensive doctrines prevailing at any one time in a given state (PL, Introduction, p. xvi; see also Cohen 1993, p. 272 and Scheffler 1994, p. 7). Indeed, the argument for justice as fairness is not complete until the case for stability in the plural milieu has been presented (ibid., §IV. 8.4). After all, Rawls often speaks of comprehensive religious and philosophical pluralism as the natural and desirable outcome of the free exercise of practical reason under liberal institutions (see e.g. ibid., p. xvi).

VI.

Let us return to the theme of moral education. Our goal here is to emphasize the continuity with the earlier account of Theory with the addition of comprehensive doctrines as a kind of intermediate factor. Rawls’s most explicit discussion of the pedagogical importance of principle-guided institutions can be found in Lecture I of Political Liberalism: citizens, he says, must somehow know, through reflection, that their social order is an association of free and equal citizens, based on egalitarian political principles; it is not, nor can it ever be, the product of a fixed social order or transcendent religious values (ibid., §I.3.1). Citizens must be educated to see the world in a political liberal way: “Think of the principles of justice [and the political institutions established in light of them] as designed to form the social world in which our character and our conception of ourselves as persons are first acquired. These principles must give priority to those basic freedoms and opportunities in background institutions of civil society that enable us to become free and equal citizens in the first place, and to understand our role as persons with that status […] We have no prior identity before being in society” (PL, §I.7.1 italics added; see also Rawls 1999c, p. 173 and Rawls 2000, p. 333).

For Rawls, then, we are educated to an idea of ourselves as free persons, in possession of the right to pursue evolving conceptions of the good, by the principles implicit in our constitutional order; we are also educated to an idea of ourselves as citizens in possession of certain political
duties – the duties of reasonableness, let’s call them – towards our fellows. Indeed, in this first lecture, we encounter nearly all of the Hegelian themes of Political Liberalism (and of Rawls’s entire oeuvre): the pedagogical function of political institutions, both in terms of our obligations as citizens and our self-perception; the explicitly social-political character of our moral and intellectual development; and the role of political institutions, when guided by the principles of justice, as the essential source of our freedom.

According to Rawls, the moral ideal of citizenship – the genuine desire to abide by the principle of justice and to contribute to the society based on them – is merely one of many possible manifestations of our nature. Human nature is indeed “permissive”: it can develop in multiple, potentially contradictory ways (PL, §II.8.1; see also ibid., §VII.5). Unlike Kant’s rigidly defined account of human nature, wherein citizenship is the essential and ideal expression of our nature, Rawls’s ideal of citizenship must be understood as one possible expression of that nature in light of certain institutional circumstances (Rawls 2000, p. 218 and p. 299). Without recourse to some publicly recognized (i.e. institutionalized) standard of equality it is unlikely that this political conception of the person – of the citizen – will develop (PL, IV.7.3; see also Rawls 2001, §55.2).

Says Rawls in this vein: “Given certain assumptions specifying a reasonable human psychology and the normal conditions of human life, those who grow up under just basic institutions acquire a sense of justice and a reasoned allegiance to those institutions sufficient to render them stable […] Citizens’ sense of justice, given their traits of character and interests as formed by living under a just basic structure, is strong enough to resist the normal tendencies to injustice” (PL, §IV.2.2). This is another way of describing a uniquely political form of socialization: “Education should prepare [citizens] to be fully cooperating members of society and enable them to be self-supporting; it should also encourage the political virtues so that they want to honour the fair terms of social cooperation in their relations with the rest of society” (ibid., §V.6.3).

Indeed, the sense of justice, which Rawls describes as an “outgrowth” of institutional arrangements, leads to the acceptance of – and allegiance to – our particular institutions, as well as the principles responsible for guiding them. We are also committed to the maintenance and reform of said institutions – to the “natural duty” to advance similarly just institutional arrangements (TJ, §72 and PL, §IV.1.3; cf. Hegel 1991, §75). The sense of justice, in other
words, is both the capacity and the desire to act from a moral point of view, as defined by our shared principles of justice; it thus renders our community stable despite its inevitable pluralism (ibid., §76; see also PL, §II.1.1). Indeed, it is the central claim of Political Liberalism that the Rawlsian sense of justice is perfectly compatible with a vast multiplicity of religious and philosophical worldviews: “History tells us of a plurality of not unreasonable comprehensive doctrines” (PL, §IV.1.6; see also Rawls 2001, §11.3 and §55). For Rawls, this internal balancing act represents the pinnacle of our moral maturity. This commitment to principles is also what distinguishes the stable political community from a (mere) modus vivendi where citizens’ commitment to the basic structure is contingent on its satisfying the (comparatively narrow) interests of their particular group (ibid., §IV.3.4; see also Rawls 2001, §55.5 and §57.3).

A common allegiance to the principles of justice is thus a testament to our unanimous desire to live together with these fellows by institutionalizing fair terms of cooperation (PL, § IV.5.5 and Rawls 1999d, p. 106). The stable political community, in other words, is “a collective achievement requiring the cooperation of all” (TJ, §79). Indeed, “the preceding account of the development of morality supposes that affection for particular persons plays an essential part in the acquisition of morality” (ibid., §73). In this way, the citizens of a well-ordered society of justice as fairness do share common ends: “we may speak of the mutual good of mutual justice” (PL, §V.8.2; see also ibid., §IV.3.4, Rawls 2001, §60.2 and Cohen 1993, p. 274). Of course, there are limits to this emphasis on particularity: “While every citizen is a friend to some citizens, no citizen is a friend to all. But their common allegiance to justice provides a unified perspective from which they can adjudicate their differences” (TJ, §72).

Citizens, in other words, are not bound directly to one another – as they are in a family, say, or in smaller associations – but to their shared institutional system based as it is on deliberate, well-known justice principles (and the concomitant overlapping consensus). What we are describing, here (and below), is what Rawls calls “the good of political society” (PL, §V.7.1). Indeed, as we shall see in detail in Chapter 3, the process of creating and maintaining just institutions is often considered by a people to be “one of the significant achievements of their history” (ibid., §V.7.4). This is precisely because the culmination of this process – justice as fairness is the outcome of historical processes – is the product of conscientious political agents: for Rawls, as for Hegel, “the safety of democratic liberties requires the active participation of citizens who
possess the political virtues needed to maintain a constitutional regime” (ibid., §V.7.5; see also Taylor 1979, p. 89 and Cohen 1993, p. 275).

Like Hegel, Rawls contrasts this image of dedicated political labour in the service of institutional vitality – in the service of the common good – with the notion of private society. In fact, Rawls makes specific reference to Hegel in this vein, acknowledging their common goal of moving beyond the society of private, self-interested persons for whom public labour is a burden (ibid., §79 and PL, §VII.10; see also ibid., §74 and §76, PL, §V.7, Taylor 1979, p. 69 and Schwarzenbach 1991, p. 559). The essential contrast is this: whereas private persons regard public institutions (i.e. rights protection) as the efficient means to interest fulfilment, citizens “value their common institutions and activities as good in themselves” (ibid. italics added). In order to understand why this must be so, we must briefly return to a prominent theme of Chapter 1: there, we emphasized the two principles of justice as the most rational choice for participants (fundamentally) interested in the development of their moral powers, including the pursuit (and revision) of their particular conception(s) of the good (ibid., §11 – 12). Hence the adoption of the first principle, which institutionalizes the equal liberties in the basic structure, and the difference principle, which guarantees the resources necessary for the legitimate exercise and development of the moral powers.

And yet the nature of human potentiality is such, says Rawls, that no person can possibly realize all of their latent potentialities (regardless of their resource bundle) (ibid., §79). This is precisely to emphasize the importance of choice – an essential moral power; we must, that is, make the conscious choice to develop (or pursue) this or that ability (or interest). And our fellows are faced with the same dilemma: each is forced to make a choice about the direction in which the moral powers are exercised. The important point is this: in light of the limits of individual potentiality, citizens come to regard the moral development of others as an essential

---

42 Of course, this idea has Kantian roots (and Rawls acknowledges such in a footnote in this section). In his “Idea for Universal History,” for instance, Kant says that the individual would have to live for a vast length of time in order to make proper use of his rational faculties, and so development manifests itself at the level of the species and not the individual. But Kant is here describing the accumulation of knowledge among the entire human race over the course of its whole history (Reiss 1970, p. 42). Rawls’s emphasis, by contrast, rests on the institutional evolution of this or that particular (plural) political community. What we shall ultimately find, then, is that Rawls seizes upon a Kantian insight but gives it a Hegelian twist.
aspect of their own good. The genuine desire for mutual benefit (i.e. equal liberty) is thus an essential component of the moral psychology of the citizen (PL, §II.1.4).

Says Rawls in this vein: “We need one another as partners in ways of life that are engaged in for their own sake, and the successes and enjoyments of others are necessary for and complementary to our good […] When men are secure in the enjoyment of the exercise of their powers, they are disposed to appreciate the perfection of others” (TJ, §79 italics added; see also PL, §V.7.4). Again, Rawls: “It is through social union founded upon the needs and potentialities of its members that each person can participate in the total sum of the realized natural assets of others […] members recognize the good of each as an element in the complete activity the whole scheme of which is consented to and gives pleasure to all” (ibid.).\(^43\) The vitality – indeed, the survival – of the political community thus depends on our association with others whose moral development has unfolded along a different path (PL, Introduction, p. xxv – xxvi, §I.1.3, §VII.5 and §VIII.3). Significant pluralism, in other words, is the essential precondition of the society of justice as fairness; Political Liberalism’s extension of this idea to a multiplicity of competing (though reasonable) comprehensive doctrines seems easily reconcilable, as long as we have common principles – and a common language – which allow for efficacious shared political labour (PL, §VI.2.2; see also Dombrowski 2001, p. 39).\(^44\)

Rawls therefore describes such as community – the well-ordered community of justice as fairness – as “a social union of social unions” (ibid.; see also PL, §VIII.6). Our shared political institutions, in other words, recognize and protect the determinate individuality of persons as members of this or that “subordinate” community. This is their instrumental function: “the basic structure secures the free internal life of the various communities of interests in which persons and groups seek to achieve the ends and excellences to which they are drawn” (ibid., §82). These particular communities are “subordinate” precisely because they are made possible by

---

\(^{43}\) Think, for instance, of members of an infield contributing to an artful double play.

\(^{44}\) In the Introduction to Political Liberalism, Rawls poses the following question, which regards as central to the aim of that work: “How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious philosophical and moral doctrines” (PL, Introduction p. xviii italics added)? It is interesting to note, here, that the necessity of (reasonable) pluralism was present in Rawls’s work before considerations of (institutional) possibility (and stability) were incorporated into the theory of justice as fairness.
shared labour in the service of institutional preservation and reformation (this is the first social union of a “social union of social unions”). And yet, in another sense, these sub-communities are on an equal footing with the political union precisely because the society of justice as fairness requires by definition significant pluralism: “Persons need one another since it is only in active cooperation with others that one’s powers reach fruition. Only in a social union is the individual complete” (ibid., §79). This is what we mean when we speak of the intrinsic value of our shared institutions: they are an essential part of our system of ends – of our good. Again, Rawls: “The collective activity of justice is the pre-eminent form of human flourishing […] it is by maintaining these arrangements that persons best express their nature and achieve the widest regulative excellences of which each is capable” (ibid.).

For Rawls, the ultimate realization of our moral powers is the desire to contribute to the society characterized by self-chosen principles of justice (PL, §II.6.1; cf. Hegel 1991, §75).

Of course, we have already examined Hegel’s emphasis on the importance of community and the deep social rootedness of individuals. Says Taylor in this vein (and at this point this passage should be evocative of Rawls’s account as well): “Man has to come to a vision of himself as part of a larger life […] what is most important for man can only be attained in relation to the public life of the community, not in the private self-definition of the alienated individual” (Taylor 1979, p. 73 and p. 93; see also ibid., p. 80). Indeed, as we have just seen, Rawls self-consciously identifies with the Hegelian project of overcoming the atomism characteristic of modern society. And like Hegel, Rawls regards deep pluralism as a precondition of – not an impediment to – stable unity. We have already seen why this is so for Rawls: individual fulfilment requires that we give free rein to the unique pursuits of our fellows; their success, as individuals engaged in the project of self-development, is a crucial indicator of our own well-being and of the stability

45 As in Chapter 1, Rawls’s reference to “human flourishing” raises important questions about the internal coherence of his thought (particularly in light of his later distinction between political and comprehensive doctrines). In other words, if Rawls abides by his own political-comprehensive distinction, doesn’t he have to abandon any talk of the “good” of community? We take up this question in detail in Chapters 4 and 5.

46 It is worth reiterating, here, the contrast to Kant: for Hegel and Rawls the state is not a neutral “umpire” between competing interests as it was for Kant; indeed, on the Kantian view, the state constitutes a (mere) modus vivendi: a non-moral consensus based on the convergence of self-interest (i.e. negative freedom). Instead, the Hegelian and Rawlsian state is ultimately characterized by shared ends.
and vitality of our community. Let us return briefly to Hegel in order to draw out the important similarity of his account and Rawls’s account outlined in this section.

According to Hegel, then, in modern times a viable political society requires that differentiation be acknowledged (Cohen 1993, p. 276). Hence Hegel’s emphasis on the mediating role of corporations and estates; the state, that is, must receive the assent (via voting estates) of each class or group within society (Hegel 1991, §308). This is, in part, a logistical argument: because of the size and complexity of the modern state, universal (i.e. total, direct) participation is no longer a feasible ideal; direct democracy requires an underlying homogeneity – a robust ground of agreement – that is no longer present in the modern context. The failure to recognize this fact can have destructive consequences; this was the cause of the post-Revolutionary Terror: in attempting to institutionalize the universal will, it sought the destruction of all manifestations of particularity (Hegel 1977, §590 - §593; see also Hegel 1991, §5 and Wokler 1998, p. 37 – 38). In this way, the Jacobins attempted a return to the unity of Greek polis despite the historical emergence of differentiated classes and communities.

The modern state must also recognize individual subjectivity, which also emerges historically through revolutions of consciousness, and the concomitant self-assertion this entails; not all citizens are willing to give themselves wholly to public initiatives, and so Hegel’s political theory (as we have already seen) reserves a prominent place for the pursuit of private interest (Hegel 1991, §182). This aspect of Hegel’s argument goes beyond mere logistics: this is another way of describing the state’s necessary acceptance, and protection, of determinate particularity. In other words, the mutual recognition structure of Spirit requires “inescapable differences of culture, values and modes of life” (Taylor 1979, p. 110). Indeed, Hegel clearly saw that the forced homogenization of citizens – through some kind of nationalist ideology, which attempts to transcend the differences between classes and groups – results in either violence (i.e. Terror) or alienation (Hegel 1964, p. 249; see also Avineri 1972, ch. 3, Avineri 1970b, p. 109 – 116 and Taylor 1979, p. 117). The only stable foundation of the community, in other words, is institutional.

If, for Hegel, significant pluralism happens to be a distinguishing feature of modernity it must be preserved as ultimately rational – as serving an ethical purpose (Cohen 1993, p. 288). We must expose the hidden rationality behind the seemingly arbitrary emergence of difference within the
national community. The rational kernel seems to be this: each class contributes something different to the community, and out of each (particular) contribution an idea of the common good emerges. Says Hegel in this vein: “If the Estates hold their assemblies in public [as they should, on Hegel’s view], there is a much more lively attitude towards the state than one where Estates have no assembly” (Hegel 1991, §315A; see also §316 – §318; see also Hegel 1964, p. 262). And in his Lectures on Hegel, Rawls emphasizes the “educative role of political life,” fashioned as it is by deliberation among diverse classes and groups (Rawls 2000, p. 358). Of course, Rawls is involved in a similar endeavour; this is captured by the concept of “a social union of social unions.” We are also reminded, here, of the Introduction of Political Liberalism: the pluralism that emerges out of the Reformation is what makes possible (after time and reflection) the well-ordered society of justice as fairness (PL, Introduction, p. xxv; see also Rawls 1999b, §1.1.1).

Indeed, what we have tried to show in this chapter is that Rawls’s ideal society, even at the time of A Theory of Justice, is a differentiated totality very much in the tradition of Hegel’s Philosophy of Right.

Conclusion.

In the end, Rawls’s account of moral learning – of the gradual emergence of proper moral sentiments – is framed in terms of the development of “our innate intellectual and emotional capacities. Once the powers of understanding are mature and persons come to recognize their place in society, they appreciate the mutual benefits of establishing fair terms of social cooperation” (TJ, §69; cf. Hegel 1991, §260). In light of our previous discussion, the Hegelian echoes in such a view are obvious: citizens come to realize (after time and reflection) the mutual advantage characteristic of just social cooperation, and become committed to the preservation of the shared milieu in question. Says Rawls in this vein: “As citizens come to appreciate what a liberal conception achieves, they acquire an allegiance to it, an allegiance that becomes stronger over time” (Rawls 2001, §58.3). Here, Rawls is describing the mutual incorporation of institutional preservation into our pattern of final ends – or, reconciliation to our social world (TJ, §75). And if, as Taylor succinctly puts it, “the doctrine of Sittlichkeit is that morality reaches its completion in a community,” then Rawls is better understood in light of his engagement with Hegel (Taylor 1979, p. 84).
And like Hegel, Rawls takes this ideal of cooperation further, emphasizing the affective or emotional dimension of citizenship: “We have a natural sympathy with other persons and an innate susceptibility to the pleasures of fellow feeling [...] it is painful for us when our feelings are not in union with those of our fellows” (TJ, §69; cf. Chapter 1). Indeed, when Rawls discusses our inherent psychological sociability he endorses an Aristotelian (and, by extension, Hegelian) principle, which emphasizes the hospitality of the natural world to the realization of the just (i.e. rational) community (ibid., §65 and §76 - §79). Says Rawls in this vein: “This conception of justice [therefore] seems more likely than its rivals to transform our perspective on the social world and to reconcile us to the dispositions of the natural order and the conditions of human life” (ibid., §77; cf. Hegel 1991, Preface and Rawls 2000, p. 332).

Activity in the service of the common good, in other words, is an outgrowth of our nature, and is not experienced by individuals as a burden or an imposition (ibid., §75). Rawls therefore accepts the possibility of genuine unity, as long as said unity respects (and protects through institutional measures, such as the equal rights constitution) the particularity of its component parts. “The most stable conception of justice,” he says, “is congruent with our good, and rooted not in abnegation but in affirmation of the self” (§76; cf. Hegel 1991, §260). Again, and in noticeably Hegelian terms: “Just institutions allow for and encourage the diverse internal life of associations in which individuals realise their more particular aims” (ibid., §79; see also PL, §VII.4 and Rawls 2001, §26.4 - §26.5).
Chapter 3: The Rousseauian Dimensions of Justice as Fairness

“At the birth of societies, says Montesquieu, it is the chiefs of republics who make the institution, and after that it is the institutions that form the chiefs of republics.”

— Rousseau, Social Contract (1762)

This chapter investigates Jean-Jacques Rousseau’s influence on the moral and political philosophy of John Rawls. That influence is quite extensive and can be organized into two distinct thematic concerns: (1) the natural psychological need for recognition and for self-respect, both of which are most effectively satisfied by well-designed political and economic institutions and (2) the socializing (or pedagogical) function of principle-guided institutions, which determine which latent human propensities and sentiments emerge in social life. Before introducing these concerns, however, we must confront a series of obvious questions: Why the turn to Rousseau? What do we learn about Rawls’s normative philosophy by excavating its Rousseauian heritage? In this chapter, I advance the view that justice as fairness is deeply implicated in the politics of recognition. Indeed, the extent to which justice as fairness is motivated by psychological concerns – namely, the need for recognition and for self-respect – has not been fully appreciated by Rawls’s interpreters and critics (Sandel 1982, Sandel 1984, Okin 1989, Baehr 1996, Krause 2005 and Frazier 2007). And Rawls explicitly follows Rousseau’s solution to the lack of recognition – and, concomitantly, of self-respect – by institutionalizing the rights of citizenship in the basic institutional structure of society. In the end, we shall see that Rawls’s constant focus on the background institutional conditions in which social interaction takes place – and on the psychosocial impact this has on the citizens whose rights are protected by those institutions – is better understood in light of Rawls’s engagement with Rousseau. That the latter is not regarded as an important influence on Rawls is a major oversight in the history of political thought.⁴⁷

---

⁴⁷ In his well-known 1975 essay, “Justice: John Rawls v. The Tradition of Political Philosophy,” Allan Bloom connects Theory to the social contract tradition of which Rousseau is an exemplary member. But Bloom’s interpretive aim is to emphasize the insuperable distance between Rousseau and Rawls. Only the former considers the question of human nature – the question of what human nature is like, i.e. its fundamental desires and drives – which Rawls explicitly brackets out through his employment of the original position: that thought experiment
But there is more at stake here than the (mere) addition of a new chapter in the history of political thought: it is my view that when we expose the full extent of the Rousseauian heritage of justice as fairness, we get a more complete, nuanced – and, in my view, attractive – image of the moral and political philosophy of Rawls. As we shall see, the claim that Rawls’s thought is best characterized as a form of asocial individualism – so resounding and so frequent in canon of contemporary political philosophy – is simply not tenable (MacIntyre 1981, Sandel 1982, Buchanan 1989, MacIntyre 1989 and Taylor 1989). Such a view represents a serious misinterpretation of Rawls’s philosophy. Indeed, citizenship in the Rawlsian milieu is more active and more demanding than generally thought: the practice of citizenship entails active participation in the interpretation, reformulation and application of shared justice principles; here; we are moving beyond the narrow Kantian rendition of reasonableness to my own more demanding account of robust reasonableness. This transition becomes much clearer, I think, when we put Rawls in conversation with Rousseau. And so, Section III below raises important questions about the nature of civic virtue, political socialization and stability in a just, well-ordered political regime: here, the chapter highlights the civic republican ethic embedded, though generally overlooked, in Rawls’s *A Theory of Justice* and in his *Political Liberalism*. And as we shall see in Chapter 4, this republican ethos recurs in Rawls’s treatise on international relations, *The Law of Peoples*.

Before we proceed, though, an important methodological point is in order. As we shall see – and this was also the case in Chapter 1 – much of Rawls’s analysis of Rousseau (as outlined in the

“banishes nature from human and political things” precisely because it stops short of plunging decision-makers into the state of nature, in which self-preservation is the central imperative (and thus the only legitimate and stable basis of political obligation). According to Bloom, the social contract tradition explicitly requires fear as the impetus for entering civil society: the basis of egalitarian political rights must be, on this interpretation, the permanent fear of, and equal susceptibility to, violent death. Rawls balks at this lowest common denominator brand of egalitarianism: he is searching for a different basis of equality. But in doing so – by rejecting the low but solid wisdom of Hobbesian social science, of the view of man as a survival machine – Rawls forfeits a compelling justificatory basis for civil society and for adhering to its founding contract. For Bloom, this is deeply problematic: “If there is to be a political philosophy, man must have a nature, and it must be knowable […] Nature is the permanent standard; what the good man and the good society are depends on human nature.” Against Bloom, I suggest that Rawls does find an alternative basis for his egalitarian political society: the passion of proper self-respect, as well as the acutely painful experience of denied self-respect. As we shall see, human nature is such that individuals are desirous of a certain kind of recognition – recognition *as equals* – from the institutions that coerce them and from the fellows with whom they are engaged in social cooperation.
Lectures on the History of Political Philosophy) is decidedly unconventional. Indeed, according to the emphases and logic of Rawls’s political Lectures, Rousseau’s oeuvre belongs in the classical republican tradition, which endorses political participation as an important good among a multiplicity of equal competing goods, and not in the civic humanist tradition, which regards political activity as the privileged locus of the good life (PL, §V.7). It therefore seems to be the case that Rawls’s heterodox or idiosyncratic interpretation of Rousseau – the “Rawlsianization” of Rousseau – goes too far: if committed citizenship in a small, sovereign civic community is a comprehensive good (in the Rawlsian sense) for Rousseau – and, of course, this is the general consensus among interpreters in the history of political thought – it is difficult to reconcile this popular image of Rousseau with the presentation of his thought in Rawls’s Lectures.

The simplest explanation is this: that Rawls’s interpretation (like any interpretation) is a partisan one, and that certain aspects of Rousseau’s thought are privileged – and others marginalized – in order to preserve the (supposed) continuity between Rousseau’s Social Contract and Rawls’s philosophically agnostic political liberalism (i.e. with his refusal to endorse a universal image of the good life). If Rousseau is an important precursor to political liberalism – and Rawls often insists that he is (see highlighted passages below) – his thought must be reconcilable with the fundamental principle of that doctrine. Many readers of Rawls’s Lectures will therefore confront an unfamiliar Rousseau: one that has much in common with liberal tradition, as opposed to the

48 We find a representative passage in Habermas The Structural Transformation of the Public Sphere, 97: “The social contract demanded self-surrender without reservation; the homme was absorbed by the citoyen. Rousseau projected the unburgeois idea of an intrusively political society in which the autonomous private sphere, that is, civil society emancipated from the state, had no place […] Consequently, the general will did not emerge from the competition of private interests; such a volonté de tous would correspond to the liberal model presupposing the private autonomy eliminated in the model of the contrat social.” See also Habermas, The Inclusion of the Other, §7.1.2. As we shall see below, Rawls rejects certain aspects of this (civic humanist) account: the Lectures emphasize the permanence of private interest – even in the society of the contrat social – and therefore endorse a fundamentally different image of proper public reason (i.e. articulation of, and adherence to, the general will). Whereas, on the Habermasian view – which, again, is representative of a long tradition of interpreting Rousseau along perfectionist patriot lines – the general will is “anchored in the hearts of citizens” and therefore characterized by its effortless spontaneity, the Rawlsian view claims that the accurate articulation of the general will (in the context of Rousseau’s contrat social) requires conscientious deliberation, and even toil. See e.g. Rawls, Lectures on the History of Political Philosophy, 99.
conventional image of Rousseau as a thinker deeply critical of liberal, bourgeois values and institutions. 49

Ultimately, though, our interest in this chapter is not to arbitrate this interpretive controversy, even though Rawls has interesting and compelling reasons (outlined below) for advocating the benign view of *amour-propre* – a view shared by other prominent interpreters of Rousseau (Dent, 1989, Neuhouser 1993, Neuhouser 2008, Cohen 2010 and Kolodny 2010). Our purpose, to reiterate, is not to defend Rawls’s interpretation of Rousseau as the best interpretation. Rather, our interest, here, is in demonstrating the impact of Rawls’s view of Rousseau on his own moral and political thought. We must therefore accept the following interpretive stance: that the Lectures represent Rawls’s genuine (unique) understanding of Rousseau’s thought – there is simply no basis for claiming that Rawls’s Lectures purposefully or consciously distort the texts in question in order to preserve their reconcilability with the (political) liberal tradition – and that this account of Rousseau is what informs Rawls’s own normative philosophy (see e.g. Rawls 2007, p. 104). We are less interested in arbitrating these interpretive controversies than we are in outlining the impact of Rousseau’s thought on Rawls’s. In the end, we shall find that Rawls’s work is deeply connected to some fundamentally republican (i.e. Rousseauian) themes, including our natural sympathy with other persons, our innate susceptibility to the pleasures of fellow feeling and, most importantly, an emphasis on properly designed political institutions as the essential source of individual (moral and civic) freedom.

Let us turn now to a brief summary of the chapter’s two main purposes: (1) an account of what Rawls’s Lectures say about Rousseau (Section II) and (2) an analysis of how Rawls’s engagement with Rousseau informs his own political project (Section III). I begin, here, with (2): as mentioned above, a central claim of this chapter is that justice as fairness is importantly

---

49 This part of Rousseau – Rousseau, the ancient, let’s say – is completely absent from Rawls’s Lectures. Rawls therefore fails to ask the following question: how can Rousseau be the author of both Books I and II of the Social Contract – liberal texts in certain essential respects, as I hope to show below – and of the Considerations of the Government of Poland, with its celebrations of ethnic particularism, cultural unity and muscular nationalism? The answer, of course, is that this communitarian political sociology is Rousseau’s pessimistic reaction to what he regards as the otherwise intractable conflict between self-interest and the common good: bourgeois self-interest and religious pluralism are permanent threats to the freedom and equality of citizens, and so liberal institutions must be supplemented by a vigorous patriotic morality. But, again, Rawls does not see this tension: as we shall see, on Rawls’s interpretation the basis of social unity in a Rousseauian political world is a set of institutions conducive to the freedom and self-development (not the patriotic identification or the deep sense of belonging) of citizens.
psychological. Rawls is deeply interested in the mental states of citizens, and recognizes the political salience of socially destructive psychological sentiments, such as envy and shame. Such feelings are often associated with inequalities of political and social status, and “with the evils of status, we are much closer to Rousseau’s concerns” (ibid., p. 245).\footnote{See also Judith Shklar, Men and Citizens, 43: “Until adolescence [Rousseau] lived in perfect equality and affection with his elders. Apprenticeship had been a terrifying shock. He could now only see the privileges and freedom of the master, and the deprivation and enslavement that he himself had to endure. Envy began to consume him, and from then on he knew why all apprentices are, and must be, scoundrels. Impotence and resentment can make them nothing else.”}

It follows that a constitutive aim of justice as fairness is to prevent the emergence of such sentiments – this, I shall argue, is the core of the Rousseauian interpretation (or dimension) of Rawls’s political doctrine. Indeed, like Rousseau, Rawls is deeply involved in the politics of recognition: from Rousseau, Rawls takes the human proclivity for interpersonal comparisons, accepts the negative consequences of unfavourable comparisons and (still following Rousseau) includes these psychological tendencies – to compare and to feel shame – as a fundamental consideration for constitutional and institutional design. It is in this spirit that Rawls claims that self-respect is “perhaps the most important primary good” (TJ, 386).

Indeed, on Rawls’s view, Rousseau’s wholly original solution to the lack of recognition – and to the concomitant lack of self-respect amongst those disadvantaged members of society – is to institutionalize political (and relative economic) equality in the basic structure of society. Rawls follows this solution, and has similarly psychological motivations for doing so. A central claim of the chapter, by extension, is that this form of equality at the highest level – at the level of citizenship – is the social basis of self-respect, Rawls’s most important primary social good (TJ, §15). Social and political conditions, in other words, are the fundamental determinant of a person’s self-respect: to be a citizen is to be secure about one’s place in society. This fundamental insight Rawls takes from Rousseau, and Rawls places it at the centre of his own normative political theory.

This is connected to the second thematic concern mentioned above: we are educated (or socialized) to an idea of equality by our egalitarian institutional milieu. Indeed, Rawls’s interpretation of the Second Discourse and of the Social Contract strongly emphasizes the highly permissive nature of the human sentiments, and the role of socio-political institutions in

---

50 See also Judith Shklar, Men and Citizens, 43: “Until adolescence [Rousseau] lived in perfect equality and affection with his elders. Apprenticeship had been a terrifying shock. He could now only see the privileges and freedom of the master, and the deprivation and enslavement that he himself had to endure. Envy began to consume him, and from then on he knew why all apprentices are, and must be, scoundrels. Impotence and resentment can make them nothing else.”
determining which of those sentiments are instantiated in social life. *Amour-propre* is therefore described by Rawls throughout the *Lectures* as (at least potentially) benign – as *reconcilable with equality*. Again, according to many commentators, this is a decidedly heterodox view: *amour-propre* is unnatural and perverse, the will to dominate (Masters 1968, Shklar 1969, Bloom 1979 and Gauthier 2006). On this view, *amour-propre* is an illegitimate form of inflamed self-regard, and does not have a benign parallel (i.e. legitimate self-respect, as Rawls insists it does, at least under the right political circumstances).\(^51\) Recalling the brief methodological discussion above, though, it is important to re-emphasize the focus of this chapter: to outline Rawls’s views on Rousseau – as opposed to trying to determine whether Rawls’s interpretation is a good interpretation – and to carry over that account into an analysis of the conspicuously Rousseauian dimensions of justice as fairness.

I.

The Kantian interpretation of justice as fairness emphasizes the principles of justice as the source of *autonomy*: autonomy, here, is *acting from principles* that express our nature as free and equal rational beings (i.e. principles chosen from behind the veil of ignorance).\(^52\) Abiding by one’s sense of justice – this is another way of describing the individual citizen’s internalization of the fairly chosen principles of justice – is the essential expression of moral autonomy (*TJ*, §79). In addition to securing the necessary conditions for autonomy, though, a central purpose of justice as fairness is to mitigate the natural human propensity to envy, which undermines the stability of the political community. Envy, on Rawls’s view, is “a [justifiable] reaction to the loss of self-respect in circumstances where it would be unreasonable to expect someone to feel differently.”

---

\(^{51}\) See e.g. Neuhouser, “Rousseau and the Human Drive for Recognition,” 22. There, the author cites Gourevitch’s translation of the *Second Discourse* in the 1997 edition of *The Discourses and Other Early Political Writings*: “Love of oneself is a natural sentiment that inclines every animal to watch over its own preservation and that, directed in man by reason and modified by pity, produces humanity and virtue. *Amour propre* is but a relative sentiment, artificial and born in society, that inclines each individual to think more highly of himself than of anyone else, inspires in men all the evils they do to one another, and is the true source of honour.”

In this vein, the *Rousseauian* interpretation (or dimension) of justice as fairness presented here emphasizes the principles of justice – and the institutions established in light of them – as the essential source *not only* of autonomy but of *self-respect too*. Indeed, the purpose of the rest of this section is to show that an essential dimension of Rawls’s project (following Rousseau) is to use principle-guided institutions to combat the socially destructive forms of envy that arise due to the *lack* of self-respect felt by society’s least advantaged members.

Of course, autonomy and self-respect are importantly related: we must regard *our* self-chosen ends as worthy of pursuit. And a lack of self-respect leads to an attenuated attachment to those ends, and to a compromised sense of ourselves as ends-setters (*TJ*, §81). And yet, in *Theory*, Rawls clearly distinguishes the two.\(^{53}\) The “first part” of justice as fairness is designed to secure the moral and political autonomy of contracting parties, while the “second part” looks at the psychological effects of chosen institutional structures (*TJ*, 47).\(^ {54}\) In fact, the rational, ends-setting individual is not *initially* conceived as subject to envy. The original position is therefore characterized by a fundamental lack of concern with interpersonal comparisons; in it, contracting parties are focussed *exclusively* on autonomy-securing institutions.\(^ {55}\) After all, justice principles ought not be chosen with reference to arbitrary preferences and dispositions, such as the propensity to envy. The capacity for autonomy, by contrast, *is* constitutive of our nature – it is neither incidental nor arbitrary (*TJ*, §80).

But, according to Rawls, the problem of envy cannot be permanently ignored in any theory of justice: such sentiments do exist *in society*, as opposed to the information-deficient original position where envy has no basis. And, for Rawls (as for Rousseau), political institutions are often the basic instigating *cause* of these sentiments.\(^ {56}\) Indeed, if the basic structure of society gives rise to feelings of pervasive envy – on account of the inequalities permitted by our

\(^{53}\) One can imagine an autonomous person lacking self-respect, just as one can imagine a self-respecting person lacking autonomy. Autonomy is not the sole basis of self-respect.

\(^{54}\) The issue of self-respect is also treated as a distinct question in Rawls, *Political Liberalism*, §VIII.6.


\(^{56}\) Of course, for both Rousseau and Rawls, well-ordered institutions are also the potential *antidote* to pervasive envy. See below.
principles of redistribution, say – this gives us reason to question those guiding principles. After all, feelings of envy may lead to mutually destructive policies and actions: “The individual who envies another is prepared to do things that make them both worse off, if only the discrepancy between them is sufficiently reduced” (*TJ*, 466).\(^57\) Envy, in other words, obscures the mutual (economic) advantageousness of social cooperation (when governed by fair principles of justice). But Rawls does not spend much time contemplating the economic or distributive consequences of widespread envy. Instead, his main concern is the *psychological* consequences of unfavourable interpersonal comparisons (*TJ*, §81).

Here, then, is the important test that characterizes this “second [psychological] stage”: How likely is this psychologically destructive form of envy in the society of justice as fairness? Not very, answers Rawls: “In a well-ordered society the need for status is met by the public recognition of just institutions […] No one is inclined to look beyond the constitutional affirmation of equality for further ways of securing his status” (*TJ*, 477). The antidote to pervasive envy is thus the *public institutional* recognition of all citizens as equals: “Self-respect depends upon and is encouraged by certain public features of basic social institutions” (*PL*, 318). A subordinate ranking in the public life of the community, by contrast, “would indeed be humiliating” and destructive of self-esteem (*TJ*, 469). And what about the psychological impact of unequal distributive shares? Again, the second principle of justice – and the institutional-distributive norms instantiated by it – is designed to prevent widespread envy: “The greater advantages of some are in return for the compensating benefits for the less favoured; and no one supposes that those who have a larger share are more deserving from a moral point of view” (*TJ*, 13).

Now, surely we have reason to be sceptical of Rawls’s response: we ought not accept (at face value) the claim that *institutional* equality can prevent the envy associated with *distributive* inequality. Although in *some* instances of public debate, we can imagine economic inequality receding into the background (and not affecting constitutional or policy outcomes), in other instances of political deliberation the self-confidence or influence of involved parties is directly

\(^{57}\) And so it is with the spite of the advantaged members of society also subject to the redistributive difference principle: “The spiteful man is willing to give up something to maintain the distance between himself and others” (*TJ*, 468).
tied to their distributive shares, regardless of norms of moral desert. But our main purpose, here, is to highlight Rawls’s recourse to an institutional-basic structure solution to the problem of socially destructive envy: indeed, a central purpose of well-ordered institutions is to lessen the impact of unfavourable economic and social status comparisons on self-respect. Again, an essential function of political institutions is to support the self-esteem of citizens. And when individual self-esteem is secure, the pleasures of community are apparent: “One who is confident in himself is not grudging in the appreciation of others” (TJ, 387). This is a testament to the Rousseauian influence on Rawls’s thought, and we shall return to this focus on the good of community in Section III below. First, though, let us now turn to the Lectures in order to better understand the precise nature of that influence.

II.

Rawls begins his lectures by emphasizing the comparatively broad scope of the Rousseau’s philosophical and political concerns: whereas Hobbes is primarily concerned with overcoming the problem of civil war, and whereas Locke is primarily concerned with justifying resistance to the Crown, Rousseau is a “critic of culture and civilization” (Rawls 2007, 192 italics added). Rousseau’s philosophical oeuvre is thus characterized by its diagnostic quality: he seeks to expose the evils of contemporary society, and to explain why certain vices have come about. Of course, on Rousseau’s view, these evils and vices are the by-product of the poorly designed political “basic framework”: “man is naturally good,” says Rawls of Rousseau’s central insight, “and it is through institutions alone that men become bad” (Rawls 2007, 192). There are two important implications to draw out of this passage: first, to assert that “man is good,” is to assert that we are not the vicious, hurtful and anti-social creatures that we so often appear to be – that human nature is reconcilable with a legitimate and stable (i.e., egalitarian) system of social and political institutions. The just, legitimate and stable regime is therefore a “realistic utopia”

58

58 Here is Rawls’s pithy definition of the realistic utopia from Rawls, Political Liberalism, lx: “We must start with the assumption that a reasonably just political society is possible, and for it to be possible, human beings must have a moral nature, not of course a perfect such nature, yet one that can understand, act on, and be sufficiently moved by a reasonable political conception of right and justice to support a society guided by its ideals and principles.” See also Rawls, The Law of Peoples, I.1.1, where the realistic utopia is defined with explicit reference to Rousseau: in order to be realistic, Rawls’s liberal conception of justice “takes people as they are,”i.e. finitely malleable and yet
Rawls’s *Lectures* therefore emphasize (repeatedly) that, on Rousseau’s view, the basic structure *can* be reformed in order to prevent the emergence of socially destructive psychological sentiments.\(^{59}\) And, second, to assert that “it is through political institutions alone that men become bad,” is to assert that institutions exercise an important – no, decisive – influence on the character of human beings: individuals take their cue from the political institutions that coerce them, and so we ought not expect a people to be anything other than what their institutions make them. In the end, Rawls’s lectures on Rousseau can be neatly divided into two main sections: Rousseau *the critic* and Rousseau *the reformer*.

Of course, the main source of the critical view is the *Second Discourse*. There, Rousseau gives his account of the history of mankind, beginning with the earliest stage of the state of nature and ending with the emergence of political authority and civil society. The crucial insight of the *Second Discourse* is this: that the present hostilities and vices of civilization are directly connected to inequalities in political power and, by extension, social status. Indeed, the social and cultural development of Rousseau’s savage leads to a new form of desire: the natural savage is no longer exclusively concerned with sustenance; he is now concerned with his *relative standing* as a member of this or that group (Rousseau 1964, 151 – 157). Rousseau is describing, here, the seemingly inevitable rise of *amour-propre*: whereas the savage is characterized by his self-sufficiency and isolation – he always “carries his entire self” with him – life in society leads to an overwhelming concern with what *others* think of us (ibid., 152 – 148). We want to see ourselves highly esteemed in the eyes of our contemporaries and, in this way, come to be dependent on their approbation (Rousseau 1975, 209). Rousseau’s point, here, is that permanent proximity generates enduring ties, and that those ties are the eventual source of sentiments of both love and jealousy; indeed, *any* form of natural difference occasions feelings of vanity, shame and envy (Rousseau 1964, 149). And as our reason, reflection and imagination become more sophisticated, we find ourselves further dissociated from, and less compassionate towards, our fellows (ibid., 132).

still capable of endorsing and acting upon egalitarian conceptions off justice, thus following Rousseau’s opening thought in the *Social Contract*.

\(^{59}\) For a contrary perspective, see Shklar, *Men and Citizens*, 2, 8 and 104.
This emergence of *amour-propre* leads us into a central interpretive controversy in Rousseau scholarship. In fact, it is a central claim of the present chapter that Rawls’s position on the matter points towards a decisive Rousseauian influence on his (Rawls’s) own normative philosophy. This controversy turns on the following question: is *amour-propre* malign by definition? Does social life necessarily lead to illegitimate and socially destructive forms of self-regard? The conventional wisdom responds in the affirmative: *amour-propre* is malign and perverse, and is meaningfully contrasted with the natural self-love (*amour de soi*) of the savage (Bloom 1997). Although *amour-propre* is the seemingly inevitable outcome of historical processes – which are themselves propelled by natural, latent human capacities – its emergence is still rightly lamented. After all, on this view, *amour-propre* manifests itself as the desire to be superior to others, to be admired by them and to dominate them. This dominating, perverse *amour-propre* is further exacerbated by economic inequality: with the emergence of agriculture and metallurgy comes the division of labour, private property and class conflict. And the tragedy at the heart of the *Second Discourse* is the transformation of *economic* forms of inequality into *political* domination. The majority of citizens are thus left in a condition of servile dependence, permanently subject to psychologically and socially destructive forms of envy and obsequiousness.

But Rawls rejects this (popular) pessimistic view, and he takes his cue from the last paragraph of the *Second Discourse*. The following two passages are therefore the essential interpretive key to Rawls’s account of Rousseau: “This [prevailing] state of society [and culture, characterized by widespread inflamed *amour-propre* and political domination] is not the original state of man; and […] it is the spirit of society alone, and the inequality it engenders, which thus change and alter all our natural inclinations” (Rousseau 1964, 180 italics added). Again: “Inequality, being almost null in the state of nature, […] becomes stable and legitimate [only] by the establishment of property and laws” (ibid.). These claims lead Rawls to adopt what would strike most Rousseau interpreters as an unconventional – indeed, *radically revisionist* – view: *amour-propre*, he says in the *Lectures*, *does* have a benign form. In Émile, for example, *amour-propre* is

---

60 See also Shklar, *Men and Citizens*, ch. 2. It is worth noting, here, that on the same page (p. 35) that Shklar describes *amour-propre* as “the desire for inequality,” she also makes reference to Émile’s seeing others “not as tools to be used, but as beings like [himself], subject to the same necessities, sensations and pains.”
potentially “cruel and malignant,” but it is also potentially “human and gentle,” associated with the “passions of beneficence and commiseration [not envy or covetousness]” (Rousseau 1979, 235). As we shall see, Rawls’s idiosyncratic interpretation of Rousseau gets carried over into his own normative philosophy: Rawls will emphasize the fact that the necessary psychological support for egalitarian institutions does exist – it is reconcilable with human nature, however permissive that nature might be – and that it is the prevailing political-institutional circumstances of society that dictate the emergence of this or that form of (malignant or benign-egalitarian) amour-propre.

In its “proper,” or benign, form, says Rawls, amour-propre is “a need which directs us to secure for ourselves equal standing along with others […] we are ready to grant the very same standing to others, and to recognize the rightful limits that their needs and rightful claims impose on us, provided – and this is essential – our equal status is accepted and made secure in social arrangements” (Rawls 2007, 199). 61 On this heterodox view, when amour-propre first emerges it is an expression of our fundamentally social nature – of the natural capacity for equality – and not a repudiation of our sociability. There is, in other words, meaningful continuity between the compassion of the savage and the proper amour-propre of the social being however temporally distant the two may be. Says Rousseau in this vein: “[Man’s] first contemplation of himself gave him his first surge of pride. Still scarcely able to grasp the concept of a hierarchical order, he saw [only] that his species ranked above all others” (Rousseau 1964, 144 italics added). 62 The clearest formulation, though, comes from Émile: there, Rousseau calls amour-propre the

61 Because this view of benign amour-propre runs so fundamentally counter to the conventional wisdom on Rousseau’s thought, it is worth noting here that Rawls’s view is in fact shared by other prominent interpreters of Rousseau. Says Dent in a strikingly similar formulation to Rawls’s: Rousseau “does not see amour-propre as inevitably taking on an aggressive and controlling character […] it directs us to secure for ourselves recognition from others and a standing in society in which we are honoured as significant beings whose needs and desires have an absolute title to be taken into account on the same footing with anyone else’s” (Dent, A Rousseau Dictionary, 35 italics added). Adds Neuhouser: amour-propre is “capable of manifesting itself in numerous ways and therefore of giving rise to an extremely diverse set of needs [including] the need to have a recognized standing among beings of one’s own kind” (Neuhouser, “Freedom, Dependence and the General Will, 378).

62 The psychological roots of this egalitarian social sentiment is the natural compassion of the savage, which is gradually dulled by the development of reason and by the rise of inflamed amour-propre: “Compassion is a natural feeling which, by moderating the activity of each individual’s self-love, contributes to the preservation of the whole species […] It supplies a [more useful] maxim: ‘Do good to yourself with as little harm to others as possible” (Ibid., 133).
distinctively human yearning “to have a position, to be a part, to count for something” (Rousseau 1979, 160).  

There is, according to Rawls’s Lectures, a crucial caveat – one that is central to the present argument. On Rawls’s view of Rousseau, this egalitarian form of sociability continues to be present in political society only as the by-product of social and institutional circumstances. Says Rousseau in this vein: “Freedom is like innocence and virtue [and, presumably, equality], in that only those who have it are aware of its value, and when they lose it that also lose their taste for it” (Rousseau 1964, 171). The competition and striving for domination characteristic of modern society are, on this view, the result of the lack of (institutional-political) recognition. Prevailing political arrangements, as Rousseau sees them, are characterized by precisely this kind of status deprivation: where despotism (i.e. institutionalized inequality) reigns, he says in his Second Discourse, “integrity and duty lose all meaning, and its slaves can have no other virtue than blind obedience […] it is here that individual become equal again, because they are nothing; […] it is here that everything reverts entirely to the law of the strongest” (Rousseau 1964, 172 italics added).

Inflamed, malignant amour-propre, in other words, is the predictable response to citizens’ perceived lack of institutional recognition, honour, status and respect. If, on the other hand, mutual recognition and status are guaranteed by well-designed institutional structures, it is possible to perceive others as sources of predictable and genuine recognition. In this spirit, Joshua Cohen draws the important distinction between the “abstract potentialities” intrinsic to human nature and the “determinate expression” of those potentialities “as a result of social circumstance” (Cohen 1997, 103). This distinction drawn by Cohen is easily reconcilable with the Rawlsian view of the possibility of benign amour-propre – a possibility essential both to Rawls’s interpretation of Rousseau and to his own normative political project. Says Rawls in this vein: “Moved by amour-propre we are ready to accept and to act on a principle of reciprocity

---

63 See also Ibid., 221: “Man is the same in all stations […] To the man who thinks, all the distinctions disappear. He sees the same passions [and] sentiments in the hod-carrier and the illustrious man.”

64 Cohen calls this the “Innocence/Virtue Thesis”: “Human beings are antecedently innocent, and while our potentialities may be realized in a vicious way, it is possible for them to be expressed in a virtuous form” (Ibid., 116). See also Melzer, The Natural Goodness of Man, Part II.
whenever our culture makes it available and intelligible to us, and society’s basic arrangements establish our secure and equal standing with others […] The particular realization of our nature depends on the culture of the society in which we live” (Rawls 2007, 199 italics added). Hence Rousseau’s advice to the Corsicans: “The fundamental law of your constitution must be equality” (Rousseau 1991, 139).

It is the responsibility of the basic structure – of society’s main political and economic institutions – to promote egalitarian political relations and to regulate potentially destructive inequalities: in this spirit, in his Social Contract Rousseau reminds us that “because the force of things tends to destroy equality […] the force of legislation should always tend to maintain it” (SC, 79). This leads Rawls to an important proclamation in light of our present purposes: “This remark of Rousseau’s is an ancestor of the first reason why, in justice as fairness, the basic structure is taken as the first subject of justice” (Rawls 2007, 234). In other words, the importance of Rousseau’s thought for Rawls is its emphasis on the fact that proper self-respect is not just a quality of individuals: it is a quality of the configuration of societies. Self-respect, in other words, ought not be contingent on the mutable opinions of individuals; instead, it is rightly assured by stable well-ordered institutions.65 Rawls explicitly attributes this insight – so important to the aim and institutional design of justice as fairness – to Rousseau.

Of course, these claims are evocative of Section I, where justice as fairness was characterized by the confluence of psychological and institutional concerns, and by the claim that psychological needs – for recognition and for self-respect – are best met by well-ordered institutions. Indeed, from Rousseau, Rawls takes away the political salience of the psychological experience of denied self-respect and the destructive envy to which this experience gives rise. Says Rawls in his Lectures: “Rousseau believes that all of us must, for our happiness, respect ourselves and maintain a lively sense of self-worth;” and so, to reiterate, Rousseau’s answer to the problem of inequality is “equality at the highest level […] citizens can moderate lower-level inequalities by general laws in order to preserve conditions of personal independence so that no one is subject to

---

65 See e.g. Rousseau, Reveries of the Solitary Walker, 127: “We want to be respected by those whom we respect, and as long as I thought well of men, I could not remain indifferent to their opinion of me. I saw that the judgments of the public are often fair, but […] this very fairness is often the work of chance, that the criteria on which men base their opinions are the fruit of their passions.”
arbitrary power, and no one experiences the wounds and indignities that arouse self-love” (Rawls 2007, 247). This distinct combination of ideas – the idea of equality at the “highest level,” of citizens as equals at this level by virtue of their shared fundamental interests and (moral) capacities, and of the political salience of self-love and the damaging psychological experience of status envy – constitutes Rousseau’s “distinctive” and “original” contribution to political philosophy and to debates about what equality means (Rawls 2007, 248).

III.

Let us shift gears now, and turn to the chapter’s second aforementioned purpose: to outline in more complete detail the Rousseauian heritage of justice as fairness. How, we ask in this vein, does Rawls’s engagement with Rousseau inform his own political project? The important theme, here, is the pedagogical one mentioned in the Introduction: Rousseau’s and Rawls’s shared belief that citizens learn principles of right from living under just institutions; and, perhaps more importantly, that it is only under such institutional circumstances that citizens acquire the motivation to act from those justice principles (recall from the Second Discourse that this is one of those latent human capacities that distinguishes us from other animals). Says Rousseau in this vein (this passage is particularly important to Rawls’s interpretation): “At the birth of societies, says Montesquieu, it is the chiefs of republics who make the institution, and after that it is the institutions that form the chiefs of republics” (SC, 69). Again: “A people’s opinions arise from its constitution” (SC, 141). In other words, institutional arrangements (when properly designed) generate the social spirit necessary to sustain them: we become attached to said institutions and principles, and are willing to work in the service of their stability and reform.66 We also become attached to those fellows whose self-understanding is centred on similar political commitments.

A multitude of individuals is thus transformed into a people. In other words, institutions change human nature: one who undertakes the founding of a people “[transforms] each individual, who by himself is a perfect and solitary whole, into a part of a larger whole from which the individual

---

receives his life and his being” (SC, 69). But this is no quick, easy process. Part of the confusion (and disappointment) over the role of the legislator in Rousseau’s political philosophy is the by-product of overestimating the efficacy of those founding moments described by Rousseau in the Social Contract. Rawls’s Lectures, for instance, stress that the society of the social compact can come about in many different ways, and that it could happen gradually over many centuries: it is clear, he says, that Rousseau “never supposes that people’s entering into an agreement of any kind could be a transition from a pre-political stage to a society whose basic institutions conformed to the requisite terms of the social compact […] The institutions that fashion a general will are designed by the law-giver who persuades the people that his authority is of a higher order […] In due course, later generations come to have and to perpetuate a general will” (Rawls 2007, 241). Political autonomy – or, the sovereignty of the general will – thus emerges over time. We need to approach the matter, in other words, from an appropriately historical perspective, for “the ideal of a just constitution is always something to be worked toward” (PL, IX.3.2). 67

Once the basic structure is established, and evolves to the point where it is rightly considered to be (reasonably, but not perfectly) just, the political community will fall into a kind of stable equilibrium: the institutions themselves will generate the support necessary for their preservation; again, this is nicely captured by Rousseau’s earlier reference to Montesquieu. Says Cohen in a similar vein: “socialization in a democratic order leads to the formation of the general will” (Cohen 1986, 280). Adds Melzer: “the moral consciousness [is] an artificial creation of the state” (Melzer 1990, 106). On this view, the question of origins recedes into the background: the legislator merely introduces the bonds of shared sentiment. The bulk of the socializing work is accomplished by the just basic structure as it evolves over time: “The state in some sense creates the citizens,” says Melzer in this spirit, “rather than the reverse” (ibid., 189).

67 This also raises interesting questions about the (constitutive) historicity of justice as fairness. Indeed, in Political Liberalism, in the context of this discussion of Rousseau’s legislator, Rawls offers the following insight: “What may cause misunderstanding is the thought that, using an abstract idea like the original position as a device of representation and the imagining the parties to understand their selection of principles in perpetuity, justice as fairness apparently supposes that citizens’ conception of justice can be fixed once and for all. This overlooks the crucial point that we are in civil society and that the political conception of justice is always subject to being checked by our reflective considered judgments.” See also Rawls, A Theory of Justice, §31.
It is precisely because of the essential malleability of human nature that Books I and II of the *Social Contract* have an optimistic quality: what Rousseau describes in them is the reconcilability of human nature with an egalitarian political milieu. The *Social Contract* therefore begins by describing the basic institutional framework of a legitimate, just and *stable* political regime, a regime in which the inflamed, malignant *amour-propre* Rousseau documents in the *Second Discourse* – the vanity, greed, the will to dominate and the sense of personal dependence – would not be present. Why, exactly, is the social world of the *Social Contract* legitimate, just and stable? Let us begin by ruling out one possible answer: that political stability requires the complete absence of *self*-interest. Indeed, there is a long tradition of Rousseau scholarship that places him squarely in the tradition of the ancients, for whom an immersive patriotism – the abolition of self-interest – is the essential precondition of political stability (see e.g. Habermas 1991, esp. p. 97).

Such a view is simply not tenable. The starting premise of Rawls’s *Lectures* is that Rousseau ought to be reclaimed, decisively, for modernity: if, as Hegel saw clearly, the quintessential mark of the modern era is the existence of interests *separate from* the interest of the community – the interests one pursues in bourgeois civil society, primarily – and if the possibility of a just and lasting political community depends on the institutional recognition – and thus, the legitimization – of these quintessentially private interests, then Rousseau *is* a philosopher of the modern era. Indeed, what Rawls shows clearly in the *Lectures* is that Rousseau’s psychology of citizenship is thoroughly modern: that he accepts a permanent but *productive* tension in the hearts of citizens between self- and the common interest. In fact, this uniquely modern emergence of *self*-interest is ultimately a cause for celebration for Rousseau: it is precisely the ineluctability of self-interest that makes virtue possible (*SC* I.8.1). One cannot privilege the general over the particular, that is, if the latter does not exist; again, this psychological tension is precisely what makes virtue – and with it, freedom – possible. This is the meaning of Rousseau’s footnote from *SC* II.3.2: “If there were no different interests, the common interest

---

68 See e.g. Hegel 1991 §265A: “‘It has often been said that the end of the state is the happiness of its citizens. This is certainly true, for if their welfare is deficient, if their subjective ends are not satisfied, and if they do not find that the state as such is the means to this satisfaction, the state itself stands on an insecure footing.’”
would scarcely be sensible since it would never encounter obstacles: everything would run by itself, and politics would cease to become an art.”

Consider, in a similar spirit, the iconic opening lines of the *Social Contract*, which are essential to Rawls’s interpretation: “I want to inquire whether in the civil order there can be some legitimate and sure rule of administration, taking men as they are, and laws as they can be: In this inquiry I shall always try to combine what right permits with what interest prescribes, so that justice and utility may not be disjointed” (SC I.1). What does it mean to say that a certain fit must be achieved between “what right permits” and “what interest prescribes?” Here, Rousseau admits – *in the first sentence* – that the stable, legitimate and happy political regime cannot rely on the kind of heroic self-sacrifice characteristic of ancient regimes. Right must be reconciled with – it must be adjusted to fit – utility; self-interest ought not be fully subsumed under the category of right. After all, human beings are permanently, ineluctably self-interested, and so the stable regime must reserve a significant place for the pursuit of self-interest. It is because Rousseau recognizes this fact – the permanence of the passion of self-interest – that he regards his reasoning in the *Social Contract* as realistic and as aimed as what is possible. In other words, the opening lines of *Social Contract* give the impression that Rousseau treats self-interest as a neutral fact about the modern world and about the composition of citizens’ sentiments.69

Of course, the claim that Rousseau is a theorist of bourgeois society is hardly novel (Orwin and Tarcov 1997). But most of the secondary literature treats the emergence of bourgeois interest as a lamentable moral tragedy: Rousseau wishes deeply that things were otherwise (Cassirer 1954; Durkheim 1960; Bloom 1997; and Wootton 2011). Why? Because the bourgeois is thought to be incapable of the wholeness that characterizes the savage in the state of nature and the patriot in the political assembly: “He who in the civil order wants to preserve the primacy of the sentiments of nature does not know what he wants. Always in contradiction with himself, always floating between his inclinations and duties, he will never be either man or citizen” (Rousseau 1979, p. 40). Commentators have been quick to latch onto this passage; in it, they see

69 Consider, here, the potentially very misleading contrast at Ibid., III.2.6 (italics mine): “In a perfect legislation, the particular or individual will should be null […], and consequently the general or sovereign will should always be dominant and the sole rule of all the others.” But Rousseau does not regard such a perfect social world as a realistic possibility; hence, his emphasis on the unity of “justice and utility,” as per the opening lines of the *Social Contract.*
the unqualified, total condemnation of bourgeois modernity. Much of the secondary literature on Rousseau thus tends to conceptualize political life as either the site of a deep, organic unity – the conscientious moulding of a distinct people by a mysterious legislator – or as the site of intractable, destructive conflict between bourgeois individuals and between various factions (Bloom 1997, Dannhauser 1997 and Kautz 1997). The solution to the human predicament, according to this conventional image of Rousseau, is either radical collectivism or atomistic individualism; that neither is a realistic possibility in modernity attests to the tragic orientation of Rousseau’s thought.

But it is Rawls’s view that we learn less from Rousseau than we otherwise would when we regard the two polarities around which his thought revolves as those of the solitary, sensitive loner and of the stout, stoic patriot. There is a more constructive and fruitful approach to Rousseau’s political philosophy, and Rawls’s Lectures show us the way in this respect. In the remainder of this section, I hope to show that the psychic possibility of self-interest is not a devastating moral tragedy for Rousseau; in fact, it is the precondition of virtue, as Rousseau himself understood that term (SC I.8.1). Neither the loner nor the patriot can rightly be described as virtuous, because the former is a slave to the inclinations and the latter is a slave to duty: virtue requires toil (SC II.11.3; see also Rawls 1971 §§3,4 and 15 and Rawls 2007, p. 234).

What, then, makes legitimacy and stability possible in the modern world? A Rousseauian social world is legitimate and stable because its political institutions are genuinely conducive to the freedom and self-development of all citizens. This interest in freedom is the interpretive key to the discussion of Rousseau that follows: “To renounce one’s freedom,” he says, “is to renounce one’s quality as man, the rights of humanity, and even its duties” (SC I.4.6). And so, in a legitimate and stable social world, shared institutions are not the expression of the sectarian interests of this or that powerful group, as Rousseau believes they are in the contemporary Europe of his day. Instead, political institutions – and public policy – ought to secure the freedom and equality of all (SC II.11.1). In other words, institutions are legitimate and stable because they secure the basic, fundamental interest in freedom (or the conditions for self-development) of those over whom they exercise coercion.

This is Rousseau’s solution to the apparent paradox of political authority – the counter-intuitive fact that freedom requires obedience to law (SC I.6.4 and II.6.7). What Rousseau shows us in
Parts I and II of the *Social Contract* is that, in the legitimate (democratic) regime, obedience is actually a form *self*-obedience: one obeys the laws one gives to oneself, laws that the kinds of persons we are – persons interested in freedom and in the conditions for self-development, for self-improvement, for the development of our faculties over time – ought to want to live under (Rousseau 1964, pp. 113-115 and 124). The general will wills what all could will precisely because of this shared interests in our own freedom and self-development. It is the fact of these shared interests that makes the general will possible in the first place; without them, political life (under the conditions of modernity) would not be possible (*SC* I.6.5 and II.1.1).

The general will therefore only wills what all citizens ought to want. Why? Because the general will *is* the common good, understood as political conditions necessary for the realization of citizens’ shared interests in freedom and self-development. The general will itself has no content, it makes no tangible or concrete suggestions about what the general will requires in practice: this must be, and can be, determined by assembled citizens (*SC* II.3.3, II.4.6 and II.11.4). Instead, the general will is a way of describing a set of institutional conditions – conditions conducive to the freedom and self-development of all citizens. The general will is therefore always “upright” because it can only be what it is – it is only one thing, the common good, and can be nothing else (*SC* II.6.10 and III.15.5).

Rousseau therefore does not lament self-interest. As we have already seen, it is an ineluctable *fact* about human nature, and so both the legitimacy and stability of any social compact depend on reserving significant space for its pursuit. But he *does* lament the manifestation of a *certain kind* of self-interest. We therefore must find a way to distinguish between healthy (or productive) and unhealthy (or unproductive) forms of self-interest. Does Rousseau provide us with the resources to draw this distinction? He does: “For while the opposition of private interests made the establishment of societies necessary, it is the agreement of these same interests which made it possible. What these different interests have in common is what forms the social bond, and if there were not some point on which all interests agree, no society could exist. Now it is solely in terms of this common interest that society ought to be governed” (*SC* II.1.1; cf. ibid., II.4.5 and II.4.7-8).

This passage is essential to the argument that that Rawls presents in his *Lectures*. In it, Rousseau advances his under-appreciated view that all persons, by virtue of their humanity, have a shared
basic interest in free self-development, in giving free reign to what Rousseau regards as the uniquely human capacity for perfectibility. We are not divided about what we want: we cannot help but want to be free, free to pursue our freely chosen ideals of self-development. Perfectibility is an essential part of who we are. And so, conflicts of interests are apparent, not necessary: there is a necessary reconcilability between the interests of individuals (SC II.3.3). Rousseau goes even further: there is also a natural symmetry between the interests of individuals and the common good of the political community. In fact, for Rousseau, individuals cannot have interests separate from the community because the community has no interests other than the fulfilment of the interests of individuals (SC I.7.5, II.11.3 and III.9.4).

The result is a shocking one: even the bourgeois is capable of wholeness. How, exactly? Because, in light of the preceding analysis, it is clear that Rousseau does not think of the particular will and the general will as two different wills or as two different forms of willing: they both make claims about how political power ought to be exercised. More fundamentally, they are both expressions of our basic interest in free self-development. But what Rousseau calls the particular will is the mistaken expression of what the general will requires in actual practice: “The constant will of all the members of the State is the general will; it is through it that they are citizens and free. When a law is proposed in the People’s assembly, what they are being asked is not exactly whether they approve the proposal or reject it, but whether it does or does not conform to the general will, which is theirs; everyone states his opinion about this by casting a ballot, and the tally of the votes yields the declaration of the general will. Therefore when the opinion contrary to my own prevails, it proves nothing more than that I made a mistake and that what I took to the general will was not. If my particular opinion had prevailed, I would have done something other than what I had willed, and it is then that I would not have been free” (SC IV.2.8; cf. ibid., II.6.10 and IV.1.6).

This is a very important passage: this aforementioned image of bourgeois wholeness is a distant ideal, for Rousseau, because certain systems of belief – capitalist greed and religious fanaticism, to name two obvious examples – have the effect of the obscuring what the actual or genuine pursuit of self-interest requires. While we share an underlying interest in free self-development, we are divided about how to achieve such a condition: our beliefs – our comprehensive doctrines, which have a decisive immediacy, a more compelling psychic presence and pull – lead
us to the mistaken view that our own good is irreconcilable with the good of others, that the
fulfilment of self-interest is a zero-sum game (SC I.7.4-7, II.3.1-3, III.4.2 and IV.1.6).

This is the illusion that Rousseau wants to expose: we ought not let our beliefs lead us into short
sightedness; we ought not let our beliefs undermine the possibility of mutually advantageous
political cooperation. It is easy to see how this error manifests itself in actual political practice:
certain citizens want to co-opt the instruments of political power in order to satisfy their own
interests, but are too short-sighted to see that, in the plural milieu, this is not a tenable approach
to the exercise of coercive power. Greedy plutocrats and religious believers with a persecuting,
converting zeal obviously come to mind here (SC III.10.1 and IV.8.6). These citizens fail to see
that the logic of sectarian voting can eventually be turned against them. And so, if we endorse
conditions that go against the general will – voting positions that undermine the institutional and
public policy conditions necessary for the free self-development of all – we must be “forced to
be free” (SC I.7.8). Rousseau’s language is inflammatory to be sure, but the theoretical point is a
relatively intuitive one.

The essential point to draw out of this section is that, according to the logic of this under-
appreciated Rousseauian argument, factional antagonism is surmountable on account of the
deeper interests that we all share. Commentators on Rousseau need no longer retreat to the
conventional loner-patriot dichotomy. Instead, the more productive, interesting and relevant
tension in Rousseau’s political philosophy is between the bourgeois-believer and the citizen,
between the pursuer of self-interest – the one who aspires to co-opt the instruments of political
power – and the potential contributor to the common good, i.e., to the institutional and social
conditions conducive to the free self-development of all. This latter tension is more interesting
and productive precisely because the two possibilities present in it are part of the same moral
universe: the tension between bourgeois-believer and citizen takes place in the political
community. It might therefore be possible to transcend this tension, and this is precisely what is
taking place in Rawls’s political philosophy, to which we turn below and (especially) in Chapter 5.

In fact, the logic of Rousseau’s argument suggests that this process – the ascendance of the
general over the particular will – should get easier over time: institutions, recall, have a tendency
to transform the worldviews and values of the individuals living under them, and so egalitarian
institutions ought to transform instinctive partiality into a will to equality (SC II.1.3). The liberal institutions described in Parts I and II of the Social Contract should have precisely this effect: the society of the general will – the society whose acts are guided by the general will, the society whose public policies are the genuine expression of the general will – should according to the logic of Rousseau’s central philosophical principle make the general will supreme in citizens’ hearts. If institutions educate, it is reasonable to believe that unconstructive, sectarian belief systems will be transformed by contact with them.

But Rousseau’s pessimism gets the better of him: the Social Contract does not end with Part II. 70 Instead, Parts III and IV go on to describe the supplementary institutions necessary to ensure the freedom and equality of citizens. In other words, Rousseau balks at the logic of his own argument: Rousseau does not believe that institutions alone are adequate; the pathologies of modern social life require more than sound institutional arrangements. What is also needed is attention to moeurs: without a vigorous patriotic morality, counsels Rousseau, no state (however legitimate in the eyes of those subject to its coercive measures) can survive (SC III.1.20, III.2.7, III.4.5, III.7.16, and III.15.3; see also Rousseau 1979, p. 40). Rousseau thus stands at the edge of modernity: he regards the flowering of private interest – although an ineradicable feature of the kinds of beings we are – as the existential threat to the political community. And so, he retreats from the precipice of modernity, to the familiar and comfortable terrain of the ancient polis where this tension between self- and public interest does not, and cannot, exist.

IV.

What Rousseau calls virtue, Rawls calls full autonomy. But the structure of both concepts is the same: virtue and full autonomy are achieved on those occasions when the particular will is subsumed under the general will. The following passage from PL captures this point: full autonomy, says Rawls, is “realized in public life by affirming the political principle of justice […] [it is] expressed by acting from the public principles of justice understood as specifying the

70 Admittedly, Rousseau’s pessimism is present in what I am calling the “optimistic” parts of SC. Cf. SC II.1.3. But, in general, the tone of the earlier parts is conspicuously different from the unrestrained pessimism of the later parts.
fair terms of cooperation” (PL, §II.6.2). We may, instinctively, want to see our own comprehensive doctrine institutionalized in the basic structure – to the benefit of its holders, and to the detriment of all others – but full autonomy is the expression of our principled commitment to securing the political and social conditions conducive to the free self-development of all citizens regardless of the content of their private beliefs or the structure of their chosen ends. Rousseau and Rawls are working in the same tradition, and their respective political philosophies, despite their many differences, share the same goals: freedom and equality.

Rawls is fully conscious of this affinity. This comes out clearly in his later discussion of public reason: “Public reason gives a view about voting on fundamental questions reminiscent of Rousseau’s *Social Contract*. He saw voting as ideally expressing our opinion as to which of the alternatives best advances the common good” (PL, 219-220; cf. SC IV.2.8). Public reason is thus the “reason of citizens”: citizens try to explain policy preferences to their fellows in ways all can endorse given, first, each parties’ privately held religious or philosophical morality and, second, and more fundamentally, each parties’ interest in free self-development (ibid. italics mine). Public reason, in other words, is the fullest, most essential expression of full autonomy: one has to learn to think and to speak with reference to the common good – understood as the institutional conditions conducive to universal free self-development – not with exclusive reference to the sectarian tenets of one’s private belief system. *This* is the most needful thing when it comes to the question of political stability. And this is precisely the psychic transformation that Rousseau regards as both essential and impossible.

But Rawls, unlike Rousseau, is best characterized by his optimism here: adopting a conspicuously Rousseauian attitude towards the question of political socialization, Rawls believes that well-ordered institutions tend to liberalize the authoritarian and sectarian tendencies of the comprehensive doctrines with which they come into contact. As we shall see in more detail in Chapter 5, in *Political Liberalism*, Rawls describes the salutary transformation of precisely those comprehensive doctrines that Rousseau ruled out as the viable foundation of justice and stability. In other words, it is clear throughout Rawls’s oeuvre – from *A Theory of Justice* all the way to *The Law of Peoples* – that he shares this fundamentally Rousseauian view: that the basic structure *constitutes individuals as citizens* and that, by extension, the *absence* of a just basic structure prevents the complete emergence of the moral sense necessary for citizenship. Indeed, justice principles effectively guide the conduct of citizens *only if* they are
publicly promulgated (i.e. institutionalized in the basic structure). Says Rawls in this vein: “Moral education has been regulated by the principles of right and justice to which [all citizens] would consent in an initial situation in which all have equal representation as moral persons” (*TJ*, 451 italics added). Again, Rawls: “Principles [of justice] regulate the practices of moral instruction in a well-ordered society” (*TJ*, 451). It follows that for Rawls (as for Rousseau) living under appropriate justice principles – which is to say, being educated by them – has a transformative effect on the social consciousness of citizens: “The acceptance of the principles of right and justice forges the bonds of civic friendship and establishes the basis of comity amidst [conflict]. Citizens are able to recognize one another’s good faith and desire for justice even though agreement may occasionally break down” (*TJ*, 454).

This important point is worth re-iterating, especially given later chapters’ emphasis on the questions of community and public reason: while Rousseau and Rawls both recognize that individual citizens will often make mistakes about what is conducive to the common good – they often let private or particular interest dictate public voting patterns, and this will lead to the breakdown of agreements in actual political practice – in the well-ordered community, the deliberative reason associated with the general will usually prevails; hence, those wrong-headed voters are “forced to be free”: to remain obedient to laws they *ought to have* prescribed for themselves given the kind of equality and impartiality constitutive of the social contract. And given the influence of Rousseau on Rawls’s thinking about public reason, we ought not be surprised to find that Rawls too has his own version of the “forced to be free” principle. Consider, for example, this often overlooked passage from *Theory*: “A person’s conscience is misguided when he seeks to impose on us conditions that violate the principle to which each would consent [in the original position]. And we can resist his plans in those ways that would be authorized when the conflict is viewed from that perspective […] there is no violation of autonomy so long as [the principles of justice] are properly followed” (*TJ*, 455).

In the end, Rawls’s well-ordered society aims to achieve the *good* of community in a way that is very much reminiscent of Rousseau. Even *Theory* has a decisively communitarian bent. Says Rawls in this representative passage: “The members of a well-ordered society have the common aim of cooperating together to realize their own and another’s nature in ways allowed by the principles of justice […] Each citizen wants everyone (including himself) to act from principles to which all would agree […] when everyone acts justly, all find satisfaction in the very same
thing” (*TJ*, 469). Would this passage seem out of place in the *Social Contract*? No, for it contains the central theme of that work: that the realization of one’s (moral) nature can only emerge, and be fully expressed, through citizenship (i.e. by acting in light of agreed upon justice principles and by cooperating in the realization, and maintenance, of just institutional structures). And so, like Rousseau, Rawls believes that the completeness – the complete good – of the individual depends upon membership in the well-ordered community.

Human nature *is* social, and this is no trivial point: Rawls is careful to emphasize that human sociability does not simply mean that “society is necessary for human life” or that “by living in a community men acquire needs and interests that prompt them to work together for mutual advantage” (*TJ*, 458). The view of our common institutions as (exclusively) instrumentally valuable is rejected; instead, our shared ends and institutions are “good in themselves” (*TJ*, 458). What, exactly, does Rawls mean by the intrinsic value of shared institutions? That it is only by living under just institutions, and by acting in light of their guiding principles, that our nature as free and equal moral persons is fully realized. *This* is the confluence of the individual and collective good: “It follows that the collective activity of justice,” says Rawls, “is the pre-eminent form of human flourishing [...] it is by maintaining these public arrangements that persons best express their nature” (*TJ*, 463).⁷¹

The well-ordered society is thus a “social good” – “a collective achievement requiring the cooperation of all” (*TJ*, 461).⁷² This takes us back, briefly, to our discussion of the institutional basis of self-respect, and to some of the earlier hesitations expressed over the capacity of institutions to assuage the envy associated with distributive and economic forms of inequality. The important point, here, is this: that in addition to the (mere) institutional recognition of our

---

⁷¹ Of course, just as the Introduction expressed reservations over the potentially comprehensive character of Rousseau’s thought, so too are we confronted with the possibility that justice as fairness, with its emphasis on politics as “the pre-eminent form of human flourishing,” *is also* a form of civic humanism despite Rawls’s protests to the contrary (see e.g. PL, §IX.3.3). If, in other words, the attempt to privilege the political (not comprehensive) character of Rousseau’s thought is a doomed interpretive endeavour – as many will insist is the case – the same may be true of Rawls’s. If this does turn out to be the case – if Rawls’s distinction between political and comprehensive doctrines ultimately breaks down – a serious re-evaluation of the internal coherence of Rawls’s normative philosophy is in order. We shall undertake this task in the remaining two chapters, and especially in Chapter 5.

⁷² See also Rawls, *Political Liberalism*, V.7.2.
status as equals, there is also an *ethos of justice* that animates citizens’ political interactions. \(^{73}\) And when, in *Political Liberalism*, Rawls mentions the emergence of such an ethos – of individuals’ striving “with evident intention” to do their part in the realization and maintenance of just institutions – he makes explicit reference to Rousseau (*PL*, 86 italics added). \(^{74}\) “The members of the community,” says Rawls, “have a common sense of justice and they are bound by ties of civic friendship […] In everyday life, the natural duties are honoured so that the more advantaged citizens do not make an ostentatious display of their higher estate calculated to demean the condition of those who have less” (*TJ*, 470). Again, Rawls: “Public attitudes of mutual respect have an essential place in maintaining a political balance and in assuring citizens of their own worth” (*TJ*, 477). Here, we have Rawls’s rendition of the Rousseauian emphasis on “seeing and being seen” and on the cultivation of civic virtue: the good of justice as fairness is “realized through citizens’ *joint activity in mutual dependence* on the appropriate actions being taken by others” (*PL*, 204 italics added).

**Conclusion.**

Why, in the end, does Rawls hold this unconventional view of Rousseau and of *amour-propre*? The benign conception of *amour-propre* is the best interpretation, according to Rawls, because it maintains the consistency and coherence of Rousseau’s wider political oeuvre. *This* is the

---

\(^{73}\) Consider this important passage from *Political Liberalism*, in which Rawls offers a comparison of his own doctrine with classical republicanism and with civic humanism: “There is no fundamental opposition [between justice as fairness and] classical republicanism […] the idea is that without a widespread participation in democratic politics by a vigorous and informed citizen body, even the most well-designed political institutions will fall into [the wrong hands]” (*PL*, p. 205). Instead, the meaningful contrast is between classical republicanism/justice as fairness (on the one hand) and civic humanism (on the other). The latter doctrine – the notion that political participation is the exclusive “privileged locus of the good life” – *is* incompatible with justice as fairness. This maps onto Rawls’s political-comprehensive distinction: whereas justice as fairness (and later, political liberalism) is a *political* doctrine – it “stays on the surface,” philosophically- and metaphysically-speaking, and so is reconcilable with other reasonable doctrines – civic humanism is a *comprehensive* doctrine – it espouses *only one vision of the good life*, the life of an engaged citizen. But this is not to say that community does not matter on the classical republican view, that the value of political participation and of egalitarian institutions is *exclusively* instrumental: as we have already seen in detail, the pursuit of shared ends is a necessary, constitutive aspect of the good of individuals; the activity of justice *is* an important dimension of human flourishing. We explore this distinction in significant detail in Chapter 5.

\(^{74}\) See also *PL*, II.1.3.
interpretive key to Rawls’s lectures on Rousseau: “the solution of the human predicament Rousseau offers in the *Social Contract* only coheres with the *Second Discourse* when we adopt the [benign] view of *amour-propre*” (Rawls 2007, 200).75 Indeed, if we reject the possibility that *amour-propre* is reconcilable with equality – if, in other words, we reject Rousseau’s view of what constitutes the natural goodness of humanity – the kind of political society depicted in the *Social Contract* is (unfortunately) utopian (not realistically utopian). There must be, in other words, psychological support for the institutional equality characteristic of Rousseau’s political philosophy: “men as they are and laws as they can be.” Otherwise, Rousseau is a “dazzling though confused and inconsistent writer” (Rawls 2007, 200). Such a view is rejected by Rawls: when social and political institutions are consistent with equality and mutual recognition, they make possible both civil and moral freedom and prevent the emergence of malignant *amour-propre*.

For Rousseau and Rawls, then, the self-respect of persons depends on the right kind of recognition – recognition as free and equal – from the political institutions that will inevitably coerce them. And, for both Rousseau and Rawls, a well-ordered institutional milieu is the only effective antidote to psychologically destructive interpersonal relations: human nature is ultimately characterized by, first, its radical permissiveness but, second, by its compatibility with egalitarian socio-political relations as long as political institutions (and the concomitant ethos of justice) teach, promote and protect the equality of persons. And what I hope to show in Chapter 4 in particular is that self-respect is a fundamental guiding concern for Rawls in both his domestic and international political theory: the goal of Rawls’s treatise on international relations, *The Law of Peoples*, is to outline the kinds of institutional arrangements capable of securing the self-respect of peoples. In other words, I hope to show that the domestic and international accounts are characterized by a signification degree of symmetry – that both accounts are motivated by Rawls’s fundamental concern with self-respect, which Rawls himself regards as a

75 There is, it must be noted, another, interpretively exogenous reason for Rawls’s holding such a view of *amour-propre*: it was Kant’s view (and of course this carries much weight with Rawls). See Kant, *Religion Within the Limits of Reason Alone*, Bk. I, Sec. I, Ak: V1:27 italics added: “We judge ourselves happy or unhappy by making comparisons with others. This is originally a desire for mere equality, to allow no one superiority about oneself, bound up with constant care lest others strive to obtain such superiority; but from this arises gradually the unjustifiable craving to win it for oneself over others.” Says Rawls of this passage: “It was not until I connected the *Second Discourse* with Kant’s remarks here that I felt I finally understood what either of them was saying. As so often, Kant is the best interpreter of Rousseau.” See also Cassirer, *The Question of Jean-Jacques Rousseau*.
quintessentially Rousseauian dimension of his political philosophy. In the end, we shall see that there is no way to disentangle the self-respect of persons from the amour-propre (i.e. the independence and concomitant self-respect) of peoples. In other words, what we get in *The Law of Peoples* is Rawls’s most complete statement of the necessary and sufficient conditions of self-respect: (1) that one receives the right kind of recognition from one’s domestic institutional milieu and (2) that one regards one’s political community as in possession of a meaningful degree of political and economic independence. The self-respect of *Theory* and the amour-propre of *The Law of Peoples* will thus be characterized by their ultimate inseparability.
Chapter 4: Bringing Robust Reasonableness Into View

This chapter marks an important transition: we shift our focus from the moral and political Lectures – as well as the Rousseauian and Hegelian features of justice as fairness outlined therein – to a reconsideration of Rawls’s own normative theory in light of its extensive, and yet underappreciated, non-Kantian heritage. This shift in focus and in emphasis is both deliberate and necessary: we are not only interested in Rawls’s interpretation of the history of political thought; we are equally, and perhaps even more, interested in Rawls’s unique contribution to that on-going dialogue. The goal of this fourth chapter is thus to synthesize the findings of the previous three and to consider the implications of the unconventional image of Rawls’s thought offered in those earlier chapters. In doing so, we get a more complete, nuanced – and, in my view, attractive – image of the moral and political philosophy of John Rawls and of Rawlsian reasonableness in particular. I call this latter notion robust reasonableness.

As we shall see, reasonableness on Rawls’s view goes well beyond mere implied or tacit consent to a given set of universally acceptable justice principles. This is the narrow Kantian view (Baumrin 1975, Darwall 1976, Pogge 1981 and Hill Jr. 2000). Instead, citizenship in the Rawlsian milieu is more active and more demanding than generally thought, and this only becomes clear when we expose the full extent of the Rousseauian and Hegelian heritage of justice as fairness. What I ultimately hope to show in the final two chapters is the conscientious demandingness of this more elaborate conception of the reasonable. This newer, thicker, robust conception of the requirements of reasonableness incorporates both the Rousseauian notion of self-respect outlined in Chapter 3 and the Hegelian notion of reconciliation outlined in Chapter 2.

Simply put, robust reasonableness captures more of what is going in Rawls’s normative thought; it is a more holistic, comprehensive and accurate image of the political morality of justice as fairness. Again, we are moving from passive acceptance of formal, universally acceptable justice principles in the original position to active participation in the interpretation, reformulation and application of shared justice principles in a deliberative democratic context (Section I). This chapter is therefore importantly indebted to the methodological discussion of Chapter 2, where the centrality of the original position in the landscape of justice as fairness was
called into question. This is a very important point worth reiterating: the kind of reciprocity modelled in the original position – what I am calling here passive consent to universal principles – is not a sufficient or faithful rendering of the kinds of demands placed upon citizens in a stable, solidaristic political community. Instead, and as we shall see in detail in our discussion of The Law of Peoples (Section II), citizens are under a moral obligation to create the conditions in which all citizens everywhere can contribute to political outcomes, if they so choose, through deliberative democratic institutions and procedures. Indeed, according to Rawls’s formulation in that late text, the legitimacy of peoples (whether liberal or decent) depends on the extent to which all citizens regard the actions and policies of their government as an adequate reflection – as genuinely representative – of the their interests, values, principles and intentions.

In the remaining two chapters, then, I hope to show that robust reasonableness is best characterized by an imperative of radical inclusiveness: the legitimate and stable community must guarantee every citizen’s equal right to participate in the political and economic affairs of the political community. It is not enough for institutions to secure the conditions for autonomous free development (through routinized, predictable redistribution, say). As we have already seen in Chapter 3, political institutions are also thought of by Rawls as the essential basis of self-respect. In order to have a meaningful sense of self-respect, that is, citizens must be recognized – and must feel themselves to be recognized – as rightful contributors to the institutional evolution of their political community. And only the political community that guarantees the conditions of self-respect is the proper object of citizens’ reconciliation. This is the confluence of the Rousseauian and Hegelian features of justice as fairness. As we shall see below, the intersection of these two under-appreciated dimensions of Rawls’s thought are on clearest display in The Law of Peoples.

I.

There are, according to the narrow Kantian view, two main virtues associated with the reasonable: first, the willingness to propose and honour fair terms of cooperation and, second, the willingness to recognize the burdens of judgment and to accept their consequences (i.e. to accept the fact of irreconcilable, and permanent, comprehensive pluralism). Reasonableness, in other words, satisfies the criterion of basic reciprocity: citizenship is about living with others –
others who subscribe to plural value systems and existential commitments – on terms that are acceptable to all persons (conceived as free and equal). The idea of public reason captures both dimensions: citizens in a democracy are to defend fundamental (i.e. constitutional) laws and policies by exclusive reference to reasons suited to the equality of citizens (see e.g. PL, §II.3.3 and §IX.6; see also Gutmann 2003, p. 185). Political liberalism is thus a doctrine best characterized by its concern with a particular conception of legitimacy in the plural milieu: it is a form of self-government in that all citizens must be thought capable of offering their seemingly implicit consent to the laws and policies that make up society’s basic structure.

But the idea of reasonableness as involving a form of implied or implicit consent is simply not satisfactory: an engaged, vigorous democratic regime is on Rawls’s view a requirement of justice – “and not simply for [the] instrumental reason” of protecting the personal liberties (or individual autonomy) (Cohen 2003, p. 87; see also Forst 2001, p. 347 and Gutmann 2003, p. 175). Indeed, it is clear that, based on Political Liberalism, Rawls endorses a distinctly deliberative (and not merely constitutional) image of democracy: according to Joshua Cohen, Rawls does regard politics as an “arena of argument, rather than a tamed competition for power, fair aggregation of interests, or expression of shared cultural [or political liberal] commitments” (ibid., p. 102; see also Hart 1989, p. 252, Habermas 1996, §7.1.2 and Dombrowski 2001, p. 9). Reasonableness is not merely the isolated contemplation (by what Kant would call “active” citizens) of what is required by the idea of reciprocity. Reasonable citizens, rather, are “ready to discuss [and to argue about] the fair terms that others propose” (PL, §II.1.1 italics added). Political autonomy (as distinguished from comprehensive accounts of moral autonomy) is therefore conceived as the “sharing with other citizens equally in the exercise of political power” (PL, p. xlii; see also §II.1.4 and §II.3.4). Says Rawls: “I am concerned only with a well-ordered constitutional democracy […] understood also as a deliberative democracy” (Rawls 1999c, p. 138). But, despite these (and various other similar) assertions, “remarkably little has been written about the

---

76 See e.g. TJ, §31: A just constitution “satisfies the principles of justice and is best calculated to lead to just and effective legislation.” See also ibid., §36 and Cohen 2003, §2.1.

77 Again, we take up this distinction in detail in Chapter 5 below.
relationship between liberalism and democracy in [Rawls’s] theory” (Gutmann 2003, p. 168).  

Our purpose, here, is to address this lacuna, with particular reference to our earlier claims about Rawls’s intellectual debts to Rousseau and to Hegel.

In this spirit, it is obvious that Rawls is not merely speaking of a democratic regime (i.e. a set of democratic institutions designed to legislate policy outcomes in line with the principle of justice and the constraints of public reason). Rather, Rawls uses “democratic” to refer to society’s prevailing ethos – to an idea of “society as a system of social cooperation among equal persons” (TJ, §59 italics added; see also Rawls 1999j, p. 246 and Forst 2001, p. 350 – 351). A democratic ethos, adds Cohen, describes a society of equal citizens “defending laws and policies by reference to reasons drawn from a conception of justice that they might reasonable expect others to endorse” (Cohen 2003, p. 103). This entails two basic, essential ideas: first, each member is understood to be entitled to be treated with equal respect regardless of their social or economic position and, second, the basis of equality is the capacity for a sense of justice; that is, we owe equal justice to those capable of understanding, and acting upon, an idea of mutually beneficial and fair cooperation (PL, §VIII.3). And so, a constitution with universal rights of participation is a requirement of justice. The contrast highlighted by Rawls is a caste or feudal society in which rights and respect are determined (and naturalized) by social tradition: “when the belief in a fixed natural order sanctioning an hierarchical society is abandoned,” says Rawls, “a tendency is set up in the direction of the two principles of justice in serial order” (TJ, §82; see also ibid., §36).

According to Rawls, then, the first principle of justice necessarily entails a principle of robust political participation. “All citizens,” he says, “are to have an equal right to take part in and determine the outcome of constitutional processes that establish the laws with which they are to comply” (TJ, §36; see also ibid., §54, PL, §I.3.4 and 1999c, p. 139 – 140). And so, the rationale for democracy is not merely instrumental: part of what it means to treat citizens as equals (as per the first principle of justice) is to give every citizen a meaningful say in the collective

---

78 This is partly explained by the fact that, together, the indexes of A Theory of Justice and Political Liberalism have only one entry (in Theory) for “democracy.”

79 As we shall see in our discussion of The Law of Peoples below, Rawls does regard certain decent, non-liberal regimes as legitimate (i.e. regimes with hierarchical or associational, which is to say non-universal, political rights).
institutional and policy choices of the community. This is precisely (to reiterate) to recognize each citizen’s capacity for a sense of justice (i.e. the ability to form, and to act upon, an idea of fair terms of social co-operation, and to shape our aims and identities in light of such an idea).

This is a very important point: at the centre of justice as fairness is an idea of deliberative democratic political participation as the most complete realization – as the essential expression – of our nature as moral persons (i.e. as possessors of an immanent sense of justice realizable under democratic institutional circumstances); hence, Rawls’s characterization of the political liberties as inviolable basic liberties (TJ, §36). Says Rawls in this vein: “It is in their public recognition and informed application of the principles of justice in their political life that citizens achieve full autonomy. Thus, full autonomy is realized by citizens when they act from the principle of justice” (PL, §II.6.1).80 This argument should be familiar from Chapter 3, where we began to unravel the relationship between full autonomy and public reason, and saw that the latter is the fullest realization of the former. But we are also confronted here by an obvious concern: it seems to be the case that this notion of full autonomy is precisely the kind of argument to which Rawls does not have recourse. Why? Because the justificatory basis of political liberalism cannot turn on the (Aristotelian or civic humanist) idea that human beings have a political nature – that human beings realize their true good, and thus live the best kind of life – by exercising their sense of justice (i.e. by actively participating in the community’s political affairs). How, then, does Rawls ultimately justify the basic status of the political liberties (as necessary for what he calls “full autonomy”)?

Rawls’s answer, I shall argue, is a fundamentally Rousseauian one (and should be familiar from Chapter 3): the public status of free and equal citizen (institutionalized in the basic structure of society) – that is, my status as someone who is rightly given a share of responsibility in the determination of the distribution of the benefits and burdens of social cooperation – is an essential guarantor of an individual’s self-respect; when, conversely, one is only an instrument of

80 See also PL, §IX.3.4: “[Habermas] and I are debating […] whether justice as fairness allows and is consistent with the [kind of discursively generated] popular sovereignty he cherishes. I have urged that it is.” We take up Habermas’s critique in more detail below and in the next chapter.
alien forces, the conditions for widespread and genuine self-respect are negated.\footnote{As we shall see below, The Law of Peoples “collectivizes” this sentiment: the genuine self-respect of peoples will depend on the extent to which their political and economic actions are the expression of their self-determination (as opposed to the effects of alien, coercive, non-representative actors and institutions).} Adds Cohen in this vein: “The idea is that others show respect for me by expressing their willingness to share responsibility on equal terms for making judgments of justice that provide supreme guidance for collective political life – not simply by recognizing me as an equal in some way, or attributing to me some equal rights regardless of the content of those rights, but as an equal with respect to making the final authoritative judgments about collective affairs” (Cohen 2003, p. 109 italics added; see also Young 1981, p. 291, Alejandro 1993, p. 77, Gutmann 2003, p. 179, Forst 2001, p. 362 and p. 366 – 368 and Forst 2007, p. 295).\footnote{The references to Forst here require a brief further comment. Indeed, it is worth noting the similarity between Rawls’s emphasis on the constitutive connection between self-respect and political participation and Forst’s recognition theory, the centrepiece of which is “the basic right […] of every member of a basic social structure to be respected as an equal participant in procedures of effective social justification” (Forst 2007, p. 295; see also ibid., p. 300 and Forst 2010, p. 717 and p. 730). Recognition, in other words, is best satisfied by maximally inclusive participatory structures of justification that are capable of changing “unjustifiable social relations;” the political thus functions as a “master dimension of justice” (ibid., p. 299 and p. 301). Persons, adds Forst, ought not be thought of as “recipients of justice;” rather, citizens should be conceived as “agents of justice” united in a normative integrated “community of responsibility” (ibid., p. 300 and Forst 2001, p. 368 – 370). Such political agency – or, “autonomy (understood in the sense of having a right to justification” – “is basic to our individual self-understanding and to our self-respect” (ibid., p. 302 italics added). With all of this, Rawls would surely agree. Of course, important differences between the two remain: Forst sees his account of deliberative democracy as a more “inclusive and dynamic” than Rawls’s liberal theory on account of the latter’s (supposed) pre-deliberative adherence to principles of justice, which function (according to Forst) as “pre-established [as opposed to discursively generated] normative guides” (Forst 2001, p. 364). This critique is the subject of a discussion below.} It is precisely in this spirit that Rawls speaks of citizenship, and the concomitant rights of equal political participation, as the highest, most important form of recognition (i.e. recognition in the basic structure) and as the essential basis of self-respect.

It is important to note, here, that self-respect is considered by Rawls to be a non-sectarian fundamental good: the importance of the basic political rights is not based on a view about the fundamental importance of political activity as the essential realization of human nature and of the human good. Instead, it is founded on the view that self-respect is, in turn, essential to a given person’s achieving their rational ends whatever those ends may be. Citizenship, recall, is the public, institutional confirmation of the value of each individual member of the community as both a moral person in possession of a sense of justice and as a rational ends-setter: “By
publicly affirming the basic liberties [including the political right to participation] citizens in a well-ordered society express their mutual respect for one another as reasonable and trustworthy, as well as their recognition of the worth all citizens attach to their way of life” (PL, §VIII.6.1).

Citizenship, in other words, fulfils two additional – and equally essential – conditions of meaningful self-respect: one regards one’s aims and ideals as worthy and, second, one believes that one is well suited to pursue them (Proudfoot 1974, p. 111 – 112, Deigh 1983, p. 229 and Forst 2007, p. 302). Again, citizenship confirms one’s good opinion of oneself (as a rational, productive ends-setter): it is a form of (institutional) recognition at the highest level. And so, “valuing the political freedom of individuals,” adds Gutmann, “morally presupposes valuing their personal freedom as well” (Gutmann 2003, p. 175; see also Shklar 1998, p. 8 and Habermas 1998, p. 260). Elsewhere, the personal and political liberties are described by both Rawls and Gutmann as “co-original” (PL, §IX.4.2 and Gutmann 2003, p. 180; see also Dombrowski 2001, p. 35 – 36 and Habermas 1996, §3.1).

This is the right moment to return to Chapter 2 and to the Hegelian notion of reconciliation that animates justice as fairness. As we saw earlier, Rawls follows Hegel in thinking of political

---

83 This notion of “co-originality” requires a brief further comment: there is, in fact, an important additional dimension of the “co-originality” thesis that characterizes Rawls’s view. On that view, part of what it means to be a free and equal member of a fair scheme of social co-operation is to be in inviolable possession of a set of liberties that includes both personal freedoms (i.e. freedom of religion and conscience, freedom of thought and speech, and freedom of association) and political freedoms (i.e. public, institutionalized recognition of one’s status as a civic equal, of one’s equal capacity to shape the institutional context in which personal freedom is exercised). In other words, both sets of liberties are constitutive aspects of the Rawlsian (free-standing, political) ideal of the person, and Rawls’s normative theory is therefore committed to securing both personal and political liberties. Adds Gutmann: “Without basic personal freedoms, on the one hand, citizens cannot truly be free to criticize their government or to stand up to a majority in the name of justice. Without basic political freedoms, on the other hand, individuals cannot be as free as possible to shape the laws, institutions, and practices within which they can make personal choices about best to live their own life” (Gutmann 2003, p. 176; see also TJ, §36). The implication of this view is important: neither the personal nor political liberties retain a prima facie absolute priority when they come into conflict, as they inevitably will (hence, our emphasis in this section on the liberal democratic character of Rawls’s theory: the political liberties are not necessarily subordinated to the personal liberties). And so, we need a way to adjudicate between such conflicts when they arise: Rawls’s theory is in principle open to arguments that would limit the basic (personal or political) liberties for all citizens for the sake of more fully realizing other basic (personal or political) liberties. So, for example, it might be possible to justify limits on the rights of political participation in order to secure freedom of conscience and religion more fully: “The priority of liberty does not exclude marginal exchanges within the system of freedom” (TJ, §37). See also PL, §VIII.12 for an example of the opposite trade-off, i.e. where the personal liberties are constrained in order to more fully realize the equal political liberties: “The mutual adjustment of the basic liberties is justified on the grounds allowed by the priority of these liberties as a family, no one of which is in itself absolute.” In other words, to reiterate, there is no prima facie standard for deciding the matter: it must be the subject of public argument, subject to appropriate discursive standards. This is precisely the role of Rawlsian public reason, which is the subject of a discussion below and in Chapter 5.
philosophy as a form of reconciliation: the task of philosophy, that is, is to show citizens that a rational, freedom-guaranteeing set of institutions already exists (Hegel 1991, Preface §4 and PL, §1.1.3 and §1.4.5). In doing so, philosophy reconcile persons to their institutional milieu. But what kind of community is the proper object of reconciliation? Of course, the political community that satisfies citizens’ basic interest in self-respect is the only legitimate object of reconciliation. And what we have just seen is that self-respect has a constitutive participatory dimension: the conditions for self-respect are undermined when there is a pervasive sense that citizens are under the coercive control of institutions over which they have no meaningful control. Self-respect therefore explicitly requires predictable, institutionalized and egalitarian access to the levers of political power; and the political community that guarantees such access generates both a sense of identification and loyalty among those subject to its coercive mechanisms. Political institutions thus serve an important ethical function: by satisfying the basic psychological needs of persons, they also generate a strong and pervasive social spirit. Self-respect and reconciliation are therefore rightly thought of as two inseparable aspects of the morality of robust reasonableness. Indeed, as we shall see in detail in our discussion of The Law of Peoples below, citizens come to recognize their institutions as the source of both freedom and self-respect, and so become committed to the preservation and enlargement of those institutions.

It is precisely in this spirit that Rawls emphasizes the intrinsic value associated with the shared (social) goal of creating and maintaining (purely political, not comprehensive) just institutions: “In the well-ordered society of justice as fairness citizens share […] the aim of ensuring that the political and social institutions are just, of giving justice to persons generally, as what citizens need for themselves and want for one another. It is not true, then, that on a liberal view citizens have no fundamental common aims. Nor is it true that that the aim of political justice is not an important part of their non-institutional or moral identity” (ibid.). Again from Theory: “Human beings have shared final ends and they value their common institutions and activities as good in themselves” (ibid.). For Rawls, then, a just society is a “genuinely social good […] and a person’s identity as citizen can be described as a constitutive attachment,” even though the basis of this constitutive attachment is a set of political (not comprehensive) values (Mulhall and Swift 2003, p. 469). Despite the political (not comprehensive) character of the Rawlsian community, working in the service of shared ends is, for Rawls, the quintessential expression of the “exercise and development of at least some of one’s higher capacities” (TI, §65; see also PL §V.7 and
Freeman 2003b, p. 290 – 296). And so, abiding by the principles of justice – and by one’s concomitant sense of justice – is “to give expression to one’s nature” and is rightly considered an essential part of “[one’s] good” (ibid., §40; see also ibid., §86 and Chapter 5 below).

Rawls therefore does have a thoughtful, compelling account of the qualitative experience of the bonds of community: if part of the challenge of political philosophy is to describe the empirical conditions in which citizens recognize their fellows as possessors of the rights of citizenship – as the proper recipients of onerous redistributive programs, for example – Rawls is deeply engaged in that endeavour and his political philosophy is better understood in light of his attempt to outline a complete vision of the good of community. Says Rawls in this vein: “Someone attaining to the more complex forms of the morality of association, as expressed say by the ideal of equal citizen, has an understanding certainly of the principles of justice. He has also developed an attachment to many particular individuals and communities” (TJ, §72; see also Miller 1995, Ch. 2, Kymlicka 1996, p. 105, Mulhall and Swift 2003, p. 478 and Nagel 2005, p. 119 – 120). In the end, then, the charge that Theory conceptualizes the individual’s relationship to the community as an instrumental, hollow one – that the individual sees the political community as a mere means to a variety of ends and not, by extension, as an intrinsically valuable cooperative venture – is no longer tenable (see e.g. Sandel 1982). Justice as fairness is ultimately characterized by the belief that doing justice – participating in the maintenance of just institutions, being an active member of a just community – is an essential part of “the human good […] that affirms our nature” (Freeman 2003b, p. 278 and p. 289; see also Proudfoot 1974, p. 114 – 115 and p. 126, Delaney 1983, p. 119, Gutmann 1985, p. 315, Alejandro 1993, p. 87 – 89 and Chapter 5 below, where we explore the implications of Rawls’s reference to a robust idea of the “the human good”).

Of course, Political Liberalism represents a partial abandonment of this (fundamentally Aristotelian) view: if Theory is characterized by the belief that justice is an intrinsic good for each person, and that the sense of justice is the most sophisticated, complete realization of our nature (and that, by extension, the pursuit of justice is rightly regulative of all other, comparatively partial goods), then Political Liberalism accepts the reality that this is only true of some citizens. The fact of reasonable pluralism is precisely the acceptance of the fact that not all citizens will regard their interest in justice – in the maintenance of just institutions – as the essential realization of their nature (even though many do retain their belief in the intrinsic good
of justice). And so, the political emphasis of *Theory* – the supposed universality of the genuine belief in, and adherence to, the principles of justice – is misplaced: *Theory* puts significant (unreasonable) demands on certain persons, particularly those whose comprehensive doctrine may assert the primacy of regulative goods other than justice (such as the belief in, and contemplation of, God and God’s will) (see e.g Dreben 2003, p. 319). The argument for stability (*for the right reasons*) therefore requires significant expansion: the intrinsic good of autonomy – or even the intrinsic good of justice – cannot do the work necessary to render a religiously and ethically plural community stable (*PL*, p. xv; see also Cohen 2003, p. 107). Stability has to be satisfied by other means.

This accounts for Rawls’s development of the overlapping consensus – the shared values implicit in society’s different prevailing (and possible) comprehensive doctrines. In fact, it is Rawls’s view that a well-ordered society will foster precisely those (reasonable) comprehensive doctrines that can accept the foundational principles and institutions of justice as fairness (see e.g. *PL*, §II.6). There will be no widely accepted comprehensive view that rejects basic liberal principles of justice (ibid., §V.6). And so, all citizens will have reasons to comply with the institutions of justice as fairness *for different reasons*. The potential threat to stability posed by the multiplicity of comprehensive views is thereby neutralized. But the implications of such a move – of Rawls’s later emphasis on the (necessity of an) overlapping consensus – are important for our purposes: it amounts to an admission that many citizens regard the pursuit of justice as subject to compromise when it conflicts with their final regulative ends. But Rawls maintains that, in the well-ordered society, such conflicts will be rare and (by extension) will not undermine the stability of the political community. Adds Freeman: “This does not deprive justice of its finality; rather it shifts finality to a more restricted domain, the political domain of public reasons” (Freeman 2003b, p. 308; see also Dreben 2003, p. 325). The purpose of Chapter 5 is to examine the philosophical coherence of this move: is limiting the regulative primacy of justice (and of autonomy) to the public political domain an option that is open to Rawls? Does it make sense to speak of the regulative priority of one’s comprehensive doctrine in only one *private* sphere of life?
II.

The decision to postpone any detailed discussion of *The Law of Peoples* was a deliberate one (and did not rest on considerations of mere chronology): it is the central claim of this section that one cannot understand that late text without a complete account of the centrality – the *fundamental importance* – of community in Rawls’s political philosophy. Indeed, the interpretive key to *The Law of Peoples* is the Rousseauian notion of *amour-propre*, which to my mind is a conspicuously communitarian concept: “This interest is a people’s proper self-respect of themselves as a people, resting on their common awareness of their trials during their history and of their culture with its accomplishments [...] this interest shows itself in a people’s insisting on receiving from other peoples a proper respect and recognition of their equality” (Rawls 1999, p. 34). Rawls’s emphasis on the autonomy of peoples is connected to this constitutive interest: just as the self-respect of *persons* depends on the institutional recognition, and protection, of the individual capacity for agency – that is, on the right kind of institutional recognition, recognition *as free and equal* – a people’s self-respect – their *amour-propre* – depends on having meaningful control over its political and economic affairs. What we get in *The Law of Peoples*, then, is an even more holistic account of the necessary and sufficient conditions of self-respect: in addition to the right kind of recognition from one’s domestic institutional milieu, another constitutive dimension of the self-respect of persons, according to Rawls’s international relations theory, is belonging to a political community with a significant degree of control over its political and economic affairs.

This emphasis on self-respect – the self-respect of persons and of peoples – helps us to make better sense of the goal of Rawls’s scheme of global distributive justice: as we shall see below, the justificatory weight for the redistributive duties that characterize the Society of Peoples rests on the fundamental importance of a people’s legitimate self-respect, which is (to reiterate) a by-product of the recognition, and concomitant non-interference, of other peoples. For Rawls, it is the value of social, cultural and political self-sufficiency – not merely the fact of economic

---

As per the argument of Chapter 3, this is an idiosyncratic rendition of this Rousseauian concept. Indeed, there is a compelling case to be made that there is no legitimate parallel to *amour-propre*: that it is (exclusively) the illegitimate desire to dominate and to degrade one’s fellow to the level of subservience. Nonetheless, the concept is used by Rawls in this way, and our explication of *The Law of Peoples* follows him in doing so. Again, see Chapter 3 for a fuller account of some of my own reservations over Rawls’s revisionist presentation of Rousseau’s thought.
interdependence or material inequality – that generates obligations of justice to non-compatriots (see e.g. Beitz 1979, Part III). Ultimately, in the Law of Peoples, a people’s *amour-propre* is of greater ethical significance than the fact of global (economic) interdependence (though the fact of robust interdependence, and concomitant inequality, *does* entail redistribution oriented towards institution-building).

What I hope to show in the remainder of this chapter, then, is that the main thrust or purpose of *The Law of Peoples* is to bring about a condition in which all peoples have a meaningful sense of their own agency, and in which all citizens regard their governments – the institutional representatives of peoples – as adequately under their control, as adequately representative of their interests and values. In fact, what we shall ultimately find is that these two imperatives are best characterized by their inextricability. For Rawls, there is no way to disentangle the self-respect of persons from the amour-propre (i.e. legitimate self-respect) of peoples: *amour-propre* – and this is, of course, merely another way of describing a condition of reconciliation to a particular institutional milieu – describes the collective sense of patriotic pride a given community has in the progressive inclusiveness of its political institutions, i.e. in the progressive enlargement of the necessary conditions for individual self-respect. 85 In other words, *the imaginative locus of amour-propre is the history of the accomplishments of robust reasonableness.*

In his treatise on international relations, Rawls seeks answers to the following questions: How should liberal peoples relate to non-liberal peoples, whether decent, outlaw or burdened states (Rawls 1999b, p. 10 and p. 19)? Should liberal peoples tolerate, and cooperate with, non-liberal peoples? If so, how far should such cooperation (i.e. *recognition*) extend? What, in other words, are the *limits* of a liberal people’s toleration of a non-liberal regime? Or, should liberal societies

---

85 Here, the notion of “a people” requires a bit further comment: although not explicitly addressed in *The Law of Peoples* – indeed, questions of immigration, migration and the resultant domestic pluralism are conspicuously absent in Rawls’s undeniably slight volume – it is clearly Rawls’s view that liberal and decent peoples have at their disposal a variety of institutional and policy options necessary to manage the intra-state pluralism that inevitably arises in 21st-century social life. The pathologies of liberal state-building *are* manageable (see e.g. Kymlicka 2008, p. 130). In other words, a people is not imagined to be an ethnically or culturally homogenous group; in fact, a people achieves this designation *only* to the extent that its citizens, who *are* animated by plural, irreconcilable values and existential commitments, regard the government as adequately representative of their interests – as a legitimate source of binding legal and moral obligation. See below.
actively seek to reshape non-liberal societies in their image, by intervening in their internal affairs (i.e. through economic sanctions or, more extremely, through military confrontations)? These are important questions for a non-ideal world: contra Kant, Rawls assumes the permanent persistence of non-liberal (or non-republican) regimes. And so, Rawls constructs the idea of a “decent” regime – to which, he would surely be willing to admit, no actual society fully corresponds (just as no actual society fully corresponds to the ideal of liberal justice) – in order to illustrate his belief in the possibility of toleration for non-liberal peoples. What, exactly, is a decent, non-liberal society? What is the basis of a liberal people’s toleration of a decent political community?

86 Adds Doyle 1997, p. 253: “Kant’s Perpetual Peace explains […] why liberal states would not, regretfully, be pacific in their relations with non-liberal states.” The notion of peoples as moral agents also presents a stark contrast with Kant. After all, in his writings on international politics, Kant dismisses the desire for cultural recognition as a mark of unsociability. Linguistic and cultural differences, he says, erect genuine barriers to the “intermingling” of nations and “occasion mutual hatred and provide pretexts for wars” (Kant 1970a, p. 114). Again: “We can scarcely help feeling a certain distaste on observing [human] activities as enacted in the great world-drama, for we find that everything as a whole is made up of […] social incompatibility, enviously competitive vanity, and insatiable desires for possession or power” (Kant 1970b, p. 42; see also ibid., p. 44 and p. 47 – 50 and Kant 1970c, p. 90 – 91). Cultural difference, that is, dulls the moral conscience; it compels us to treat others as the means to cultural self-aggrandizement rather than as ends in themselves (Hedrik 2008, p. 253; see also Wood 1990, p. 296 – 300). Indeed, the notion of a satisfied, or proper, patriotism is clearly alien to Kant’s international relations thought and this, as we saw in Chapter 1, makes it fundamentally different from Rawls’s (and from Rousseau’s too, as per the argument of Chapter 3). This is a very important point worth re-iterating: although the central aim of Kant’s international relations theory is to secure the formal consent of republican states for the treaty of perpetual peace – of course, Rawls too emphasizes the necessity of a legally constituted framework among liberal and decent peoples – a constitutive dimension of statehood – of difference – is (on Kant’s view) bellicosity (even among “civilized” states). Ultimately, it is only the teleological aspect of Kant’s thought that is able to overcome the mutual enmity produced by cultural difference. Indeed, it is characteristic of Kant’s international relations theory that it always reverts to an emphasis on the general incapacity of states to respect the autonomy – the moral personality – of other states. Despite our general moral incompetence, then, Kant can still maintain that perpetual peace is possible precisely because it is nature’s will: “How does nature guarantee that what man ought to do by the laws of freedom (but does not do) will in fact be done through nature’s compulsion? […] For I say that nature wills that this or that should happen, this does not mean that nature imposes on us a duty to do it, for duties can only be imposed by practical reason, acting without internal constraint. On the contrary, nature does it herself, whether we are willing or not” (Kant 1970a, p. 112; see also Kant 1970b, p. 41 and Kant 1970c, p. 89). In the end, Kant regards perpetual peace not as the result of human agency, but as a result of the will of a benevolent deity. Only Nature is able to overcome perpetual warfare in Kant’s system.

87 Of course, “outlaw” states (not peoples) are not to be tolerated: “An outlaw state that violates [human] rights is to be condemned and in grave cases may be subjected to forceful sanctions and even to intervention […] Outlaw states are aggressive and dangerous; all peoples are safer and more secure if such states change, or are forced to change, their ways. Otherwise, they deeply affect the international climate of power and violence” (Rawls 1999b, p. 81). See also ibid., p. 25 and p. 38: The Law of Peoples rejects the “traditional powers of sovereignty,” including, “the right to go to war in pursuit of state policies […] [and] unrestricted internal autonomy in dealing with its own people.” It is also worth noting, here, that Rawls only uses the word “people” to describe those liberal or decent (i.e. well-ordered) societies. When speaking of outlaw and burdened societies, Rawls uses terms such as “country” or “state.”
According to Rawls, a “decent hierarchical society” is characterized by the following features: (1) it is peaceful and non-expansionist; (2) it is guided by a common good conception of justice that secures the human rights of all members, including the right to life, to liberty (of thought, of conscience and of association) and to formal equality; and, most importantly, given our interest here in the demands of robust reasonableness, (3) it has a “just [and legitimate] consultation hierarchy,” which represents each segment of society, including especially women (Rawls 1999b, p. 64 – 67, p. 75 – 78 and p. 117). A decent society need not give all members equal political (as opposed to human) rights: it may be guided by a comprehensive doctrine, and so reject full political participation for some members in light of the tenets – institutionalized in the basic structure – of that sectarian doctrine. According to Rawls, such a society is not so unjust as to be indecent: although not liberal – and therefore rightly subject to public criticism by private individuals – the regime has consultative institutions and procedures, and is therefore regarded by its own citizens as representative of their interests and values; decent regimes are regarded by their own citizens as legitimate. According to Pettit, these emphases on representation and (by extension) legitimacy are essential to a proper understanding of The Law of Peoples: “The members of any well-ordered people will be party to certain shared ideas that are capable of being articulated into a theory of justice. And they will control the government that represents them” (Pettit 2006, p. 48; see also Tan 2006, p. 83; see also Wenar 2006, p. 100).

It would therefore be unreasonable for liberal regimes to refuse to cooperate with decent ones, or to insist on their becoming liberal as a prerequisite of economic or other forms of cooperation. Says Rawls in this vein: “This lack of respect may wound the self-respect of decent non-liberal

---

88 Human rights are conceived by Rawls as a “special class of rights” that specify the minimum standards of decency (i.e. the basic rights that individuals need to develop and exercise the basic capacities that are constitutive of their moral personality). They therefore function as the necessary justification for intervention: “a government’s internal autonomy is now limited” in the name of this “special class of rights” (Rawls 1999b, p. 79 and p. 114; see also Tasioulas 2002, p. 380 and Bernstein 2006, p. 282 – 286). While it may be regrettable for a (decent) regime to deny the right to vote to certain members (based on seemingly morally arbitrary criteria), it does not (for Rawls) carry the same moral weight – that is, it does not justify external coercion or intervention – as the denial of the right to life, liberty or security of person. Human rights, in other words, “set a limit to the pluralism among people” (ibid.; see also Hinsch and Stepanians 2006, p. 118). For a critical take on this view of human rights, see Beitz 2000, p. 687 – 690. This view of rights also gives Rawls recourse to a compelling response to Pogge’s aforementioned critique that Rawls ignores the interests of persons (in favour of the interests of peoples): in fact, a people is such only by virtue of protecting the basic interests of its individual members. See also Nickel 2006, p. 267 – 271 for a critical view of Rawls’s human rights “minimalism.” On Nickel’s view, human rights ought to serve more roles (than merely justifying intervention), such as guiding domestic aspirations for reform and public criticism.
people as peoples [...] and may lead to great bitterness and resentment” (ibid., p. 61). Instead, “liberal peoples must try to encourage decent peoples [by recognizing their political autonomy, and by respecting their status as equals] and not frustrate their vitality by coercively insisting that all societies be liberal” (ibid., p. 62). In other words, active coercion – the supposed necessity of economic and military intervention in the name of liberal values and institutions – is not a tenable foreign policy for liberal states. It is not the task of the Society of Peoples to enforce the full rights of democratic citizenship.

Here, both the domestic and the (uniquely) international dimensions of robust reasonableness come into full view: with regards to the domestic case, and as we saw in Section I, robust reasonableness requires the active, dedicated pursuit of inclusive (and thus representative) political institutions; again, robust reasonableness demands the installation and maintenance of a set of political institutions that are thought of by all as a legitimate source of coercion, of binding legal and moral duties. Recall Pettit’s aforementioned emphasis on the necessity of citizens being in control of their government (whether liberal or decent). But what we see here, in Rawls’s discussion of decency, is that reasonableness itself has an element of context-dependence: the secure rights of contestation (as the object of robust reasonableness and the basis of political legitimacy) ought not be thought of as exclusively the domain of individuals; political actors in decent (as opposed to liberal) regimes are also responsible for ensuring that properly constituted (and thus adequately representative) groups have institutionalized access to decision-making authorities.

In other words, with the addition of even more radical international pluralism, Rawls first refines and then expands the demands of robust reasonableness: first, participation, inclusiveness and legitimacy will require different actions and policies depending on the presence or absence of a background comprehensive doctrine that limits the scope of political institutional possibility. We are wrong to expect, in other words, that all regimes can achieve the degree of maximally inclusive participation characteristic of liberal regimes described in Section I. But what remains the case is that, in both liberal and decent regimes, legitimacy still depends upon the extent to which the state reflects the interests, values and intentions of the governed (whether the governed is thought of as discrete individuals or as the members of representative groups). And on the theme of expansion, robust reasonableness becomes even more demanding in the sense that it constitutively requires the restraint of liberal actors in their dealings with non-liberal decent
regimes. But this is not mere toleration: robust reasonableness requires respectful recognition, not begrudging acceptance. Recall, here, the aforementioned ideal of reciprocity at the heart of the Society of Peoples. This is a very important point, for this is where the full complexity of the demands of reasonableness becomes clear. After all, in a liberal regime, domestic reasonableness and international reasonableness ultimately require different things: the former requires the realization of the conditions of universal participation, while the latter counsels the practice of respectful restraint given the radical, and irreconcilable, pluralism of the contemporary global order.

On Rawls’s view, then, the international order can be just without all of its member states being fully just (in the liberal-democratic sense). And so, in the end, Rawls rejects the Kantian view that only republican (i.e. liberal) states are rightly admitted to the Society of Peoples (ibid., p. 82; see also Kant 1970b, p. 47, Kant 1970c, p. 90, Pogge 1997, p. 365 and Doyle 1997, p. 252). In fact, Kant’s insistence that all states be republican in order to qualify for admission into the pacific federation is fundamentally irreconcilable with the Rawlsian emphasis on the moral importance of respect and recognition of non-liberal peoples (Kant 1996, §60). But – and this is key – this is not a repudiation of the normative view (going all the way back to Theory) that well-orderedness – that the protection of individual rights and the pursuit of maximally inclusive political participation by individuals (not hierarchical groups) – is a condition to which all societies rightly aspire. Instead, Rawls believes that achieving full (democratic) justice should be the outcome of self-determined processes: “If a liberal constitutional democracy is, in fact, superior to other forms of society, as I believe it to be, a liberal people should have confidence in their convictions and suppose that a decent society, when offered due respect by liberal peoples, may be more likely, over time, to recognize the advantages of liberal institutions and take steps towards becoming liberal on its own” (ibid., p. 62 italics added; see also Reidy 2004, p. 305 and Nili forthcoming).

This emphasis on the importance of self-determination – and on self-determination as the essential basis or foundation of self-respect (i.e. legitimate amour-propre) – is essential to a proper understanding of The Law of Peoples: “It is surely a good for individuals and associations to be attached to their particular culture and to take part in its common public and civic life” (ibid., p. 111 italics added; see also Rawls 2001, p. 50). Again: “Peoples have a definite moral nature, which includes a certain proper pride and sense of honor; they may be proud of their
history and achievements, as a proper patriotism allows” (ibid., p. 44 italics added). Proper patriotism, in other words, is a people’s interest in “proper self-respect”: self-respect based upon a common consciousness of their historical, political and cultural achievements (ibid.). Of course, meaningful self-determination is the essential condition of its possibility. This image of a self-determined people is illuminatingly described by Pettit as a “civicity”: the members of a well-ordered people “are active in the manner characteristic of citizens, as citizens are traditionally conceived, being disposed to invigilate and interrogate those who act in the name of the collectivity” (Pettit 2006, p. 48; see also Pettit 2005, p. 158). This is a very important point: political institutions do have ethical significance precisely because they are the practical embodiment, and imaginative locus, of our collective efforts to make our institutional milieu as representative as possible – political institutions are the site of our self-determination and the basis of our self-respect. This is precisely why the universal self-determination of peoples is the “target” of global redistribution: Rawls’s duty of assistance is fulfilled once all those burdened societies achieve a meaningful degree of political autonomy and economic self-sufficiency. In other words, the purpose of global justice is to secure the autonomy and concomitant self-respect of peoples.

89 A constitutive feature of a people is that their citizens are united by “common sympathies.” Rawls turns to Mill as a source for this cultural conception of citizenship: “A portion of mankind may be said to constitute a Nationality, if they are united among themselves by common sympathies, which do not exist between them and any others – which make them cooperate with each other more willingly than with other people, desire to be under the same government, and desire that it should be government by themselves, exclusively” (Mill 1977, p. 546). While Mill considers the foundation of “common sympathies” to be a shared language and culture, it should be no surprise that the Rawlsian basis for social unity is explicitly political: it is precisely the existence of shared domestic institutions capable of securing each citizen’s sense of self-respect that functions as the essential foundation of a people’s common sympathies (PL, I.4.3, II.1.1 and V.7.2). And a people’s interest in self-respect is the desire that others recognize their unique political (and political cultural) identity. According to Rawls, then, a people’s identity is the result of a long socio-political historical process: namely, the gradual institutionalization of justice principles. “Learning from one another’s efforts and appreciating their several contributions,” he says, “human beings gradually build up systems of knowledge and belief; they work out recognized techniques for doing things and elaborate styles of feeling and expression. In these cases, the common aim is profound and complex. The essential thing is that there be a final shared end and accepted ways of advancing it which allow for the public recognition of the attainments of everyone. When this end is achieved, all find satisfaction in the very same thing; and this fact together with the complementary nature of the good of individuals affirm the tie of community” (Rawls 1971, §79; see also PL, V.7.4). This shared history is a source of authentic and, more importantly, legitimate collective pride. And so: “Leaving [one’s political community] is a grave [although not inconceivable-] step: it involves leaving the society and culture in which we have been raised, the society and culture whose language we use in speech and thought to express and understand ourselves, our aims, goals and values; the society and culture whose history, customs and conventions we depend on to find our place in our social world” (Rawls 2001, §26.5).
The important point is that self-respect cannot be disentangled from this sense of collective responsibility; and, as we have already seen, self-respect is the fundamental concern of Rawlsian peoples: this is what distinguishes them from those liberal persons discussed in *Theory*, for whom autonomy is achieved (in large part) through the institutional maintenance of relative equality (as determined by the difference principle or something resembling it). And so, the predictable (permanent) background operation of something like an international difference principle is ultimately rejected by Rawls because it is thought by him to be detrimental to the self-respect of the world’s well-ordered decent regimes. This is more than mere concern with logistics or feasibility: Rawls’s rejection of the international difference principle is not because of the absence of a world state capable of administering and enforcing a robust Poggean redistributive scheme; neither is it a concession to the desire of rich states to “hoard” their resource bundles (Pogge 1994). Instead, the distributive arrangements characteristic of the Society of Peoples are a testament to the ethical significance of self-determination, and to the value of shared (well-ordered) political institutions as the imaginative locus of political independence and, by extension, of amour-propre.

Clearly, then, Rawls does not endorse the absolution of rich states’ responsibility for the socio-economic condition of the world’s least advantaged peoples, and any suggestion that he does is implausible (see e.g. Pogge 2006). The demands of morality do extend beyond national borders: “A liberal people tries to assure reasonable justice for all its citizens and for all peoples; a liberal people can live with other peoples of like character in upholding justice and preserving peace” (*LP*, 29, 44 and 113). Peoples (not states) thus have agency and are characterized by their moral motives and capacities (*LP*, 17 and 65–66; see also Pettit 2006, p. 42 and Tan 2006, p. 79).90 As long as there are institutions and policies that negate or undermine the self-determination of peoples, there is reform work to be done: “[peoples] will formulate guidelines for setting up cooperative organizations, and will agree to standards of fairness for trade as well as to certain

90 Indeed, according to Rawls, the idea of peoples as moral agents is already implicit – although not yet fully developed – in the public international culture (and as we have already seen, for Rawls, this is decisive) (Rawls 1999a, p. 83). Recall our earlier emphasis on the ideal of reciprocity and mutual respect: not only should a government relate to its own citizens reasonably (or decently); it should also relate reasonably to other governments (ibid, p. 37). And so, as long as peoples respect human rights and adequately represent the will of its citizens, there are no grounds for coercive intervention. These are the minimum conditions of legitimacy, and are recognized, explicit features of the public international culture. Rawls’s global original position then extends our judgments about what reasonableness towards other (burdened) peoples requires: the duty of assistance.
provisions for mutual assistance. Should these cooperative organizations have unjustified distributive effects, these would have to be corrected in the basic structure of the Society of Peoples” (LP, 115). Again, “political autonomy” is the overriding goal of the Society of Peoples: economic and distributive arrangements are just insofar as they promote the freedom and equality of liberal and decent peoples, and unjust insofar as they undermine political autonomy (LP, 118).

Conclusion.

The purpose of this chapter was to distil, and to bring to the forefront, what was merely implicit in the previous three: the notion of robust reasonableness. The key idea, here, is that citizenship in the Rawlsian milieu is much more active and much more demanding than the conventional image of Rawls that emerged from the liberal-communitarian debates of the 1980s and 1990s: reasonableness is not merely the formulation of principles and laws that could (hypothetically or tacitly) be agreed to by those subject to them (in the original position, for example). Rather, robust reasonableness requires both engaged participation – via various deliberative institutions and procedures – and the maximal extension of the rights of political participation (given the contingent comprehensive limits on feasible political reform). The self-respect of individual citizens depends on the extent to which they are adequately recognized by political institutions and the extent to which said institutions adequately represent their interests, values and goals. In other words, citizens adhering to the requirements of robust reasonableness are committed to making their political institutions as inclusive and (by extension) as representative as possible. Robust reasonableness therefore leads to what I think is a fairly stringent standard of domestic legitimacy. Surely this was Rawls’s intention when he was trying to articulate the nature and institutional arrangements characteristic of a legitimate people.

What I have also tried to show is that self-respect is a fundamental guiding concern for Rawls in both his domestic and international political theory. While The Law of Peoples accepts the possibility of hierarchical (non-universal) rights of participation (in light of the tenets of some comprehensive doctrine) – this amounts, really, to a kind of multiculturalism of international relations – the fact remains that a necessary condition of the legitimacy of all regimes (whether liberal or decent) requires that citizens do not regard state coercion as a form of alien (non-
representative) interference. But the truly demanding nature of robust reasonableness comes into full view only when we important the questions and concerns of international political theory: international arrangements are reasonable only insofar as they promote the autonomy and concomitant self-respect of peoples. In other words, the demands of robust reasonableness extend beyond domestic borders: to the extent that economic-distributive-institutional arrangements undermine the self-sufficiency, and self-respect, of burdened peoples, well-ordered peoples are under an obligation of justice to reform these unequally burdensome arrangements.

And so, what I have tried to show in this chapter is that The Law of Peoples is wrongly thought of as an excessively permissive guide for the foreign policy conduct of liberal states. In fact, given the prevailing arrangements characteristic of the global economy – and the kind of institutional inertia to which such complex arrangements are usually subject – the ideal of mutually beneficial economic cooperation at the centre of The Law of Peoples strikes me as quasi-Platonic. Perhaps Rawls is more accurately criticized for his excessive optimism over the prospects of fair global economic reform? Of course, this is a question for another day. But such a view represents an almost complete reversal of the general tenor of the prevailing secondary literature on Rawls’s text (Beitz 1979, Pogge 1989, Pogge 1994, Beitz 2000, Buchanan 2000, Kuper 2000, Tan 2000, Pogge 2002 and Pogge 2006). Instead, I am suggesting that The Law of Peoples is perhaps too radical, not too conservative. If Rawls’s normative project for international relations fails, it is because of an excess of ambition, not naïveté over the empirical conditions faced by 21st-century peoples (Nili 2010).

There are two other important elements of robust reasonableness worth reiterating before we move on to Chapter 5. First, the notion that citizenship (in the plural milieu) requires the establishment of maximally inclusive participatory institutions (as the means to universal self-respect) is deeply tied to the intrinsic significance of community in the Rawlsian theoretical landscape. Citizens that have been socialized in the liberal milieu do not regard the movement towards universal inclusion as a burdensome impediment to the pursuit of self-interest. Rather, the regime of universal political participation (as the aspirational object of engaged political activity) is both a requirement of justice and the fullest expression of the good of political community. This, according to my presentation, is the Hegelian reconciliation dimension of robust reasonableness. And, second, the political community itself – the centrepiece of which is the principle-guided basic structure – is the object of citizens’ collective self-respect, their
amour-propre. Rawls clearly takes this insight from Rousseau (even though the original, malign meaning attributed to amour-propre by Rousseau is replaced with a legitimate parallel). Citizens, that is, are appropriately proud of the progressive inclusiveness of their political institutions. Again, the imaginative locus of amour-propre is the history of the accomplishments of robust reasonableness.
Chapter 5: The Width of Public Reason

We come now to the question of the neutrality of justice of fairness – rather, the neutrality of political liberalism, the later, refined version of that earlier doctrine. Why does this question of neutrality matter? It matters, because it speaks to the central aim of Rawls’s later work: to achieve political stability for the right reasons by making the state maximally inclusive; that is, by maximizing the number of citizens capable of regarding the liberal state as a legitimate source of coercion. In order to achieve such a condition, public policy and public utterances must pass the test of public reason: citizens with plural value systems and existential commitments must all be capable of offering their principled, public assent to state coercion. And so, there are various restrictions (described below) on what can reasonably be said in public and, more pertinently, on the kinds of reasons that can ultimately justify public policy. In the end, though, all of these discursive and justificatory constraints are captured by the idea of neutrality: the laws and policies of the liberal state ought not privilege the holders of a particular worldview or value system; they ought to be neutral, in other words, between the various prevailing comprehensive conceptions of the good. Political liberalism is therefore best characterized by the anti-perfectionist view that judgments about the value or desirability of a given system of ends are best left to individuals; that such judgments ought not be made by the state; and that the coercive actions of the state should not promote the flourishing of a particular system of ends. Again, an obvious corollary of this view is that the coercive policies of the state ought not make reference to – or be justified in terms of – a particular worldview or system of ends (PL, §VI.2).  

On such a view, political philosophy is no longer rightly concerned with – it must avoid on principle – the controversial questions of moral philosophy and metaphysics: as we have seen in previous chapters, the task of political philosophy on Rawls’s view is eminently practical – to articulate a basis for principled and therefore stable social unity (as opposed to a mere modus

---

91 As per the argument of earlier chapters (and of Chapter 4 in particular), the shift from Theory to Political Liberalism is perhaps best understood as a shift in theoretical emphasis from the original position – like many contemporary theorists, Rawls comes to regard hypothetical, pre-political consent as a problematic basis for the exercise of coercive political power – towards a discursive ideal of reason-giving and public deliberation.
vivendi) in the permanently plural community. Rawls’s endorsement of such stability is based on the post-Reformation historical emergence of a “new social possibility”: the possibility of tolerant (i.e. reasonable) disagreement in metaphysical areas such as religion and ethics combined with principled agreement on a political conception of justice. Peace and unity are thus secured without any citizens having to give up deeply held private beliefs. According to the view presented by Rawls in *Political Liberalism*, adds Jean Hampton, we need to “rethink what we are doing when we theorize about political life” (Hampton 1989, p. 795; see also ibid., p. 803, *PL*, §IV.5.2, Rorty 1994, p. 179 – 184 and p. 188 – 189 and Weithman 1994, p. 190 – 191 and p. 195 – 196).

But the most important claim in previous chapters was that citizenship in the Rawlsian milieu is more active and more demanding than is generally assumed: here emerges the spectre of sectarianism. We have to confront the possibility, that is, that the account of robust reasonableness developed in earlier chapters leads to a breakdown in the philosophical coherence of Rawls’s later political theory: is the image of civic virtue at the centre of Chapter 4 – which is, of course, the culmination of the previous chapters’ examination of the Rousseauian and Hegelian dimensions of Rawls’s moral and political philosophy – ultimately a violation of the kind of neutrality to which political liberalism aspires? On this critical view, the ideal of liberal neutrality is fundamentally incoherent: its supposedly anti-perfectionist emphasis on the right over the good – on the belief that the properly constituted state refuses to actively promote the private values and ends of a particular group and that, by extension, its essential purpose is the protection of the rights of its individual citizens – “smuggles in” (under the inconspicuous cover of a supposedly neutral rights-based framework) its own ideal of the good life for human beings. It thereby violates its own claim to neutrality. We must confront this important critical concern: Is robust reasonableness a violation of the kind of neutrality to which Rawls’s later political philosophy aspires?

The idea of public reason is, I think, a good entry point into this question about the neutrality of political liberalism: it gives us the opportunity to see what the public practice of robust reasonableness involves in the plural milieu, and whether its demands are ultimately dependent on a more robust conception of worthwhile ends – a fuller image of human flourishing – than Rawls’s political philosophy wants to admit into public discourse. Are the demands of public reason too onerous to be considered genuinely neutral among prevailing (and possible)
conceptions of the good? We are most clearly confronted by such worries when Rawls offers his definition of political autonomy – an ideal to which all citizens rightly aspire (even within the philosophical constraints that characterize Rawls’s political not comprehensive doctrine), and which “expresses our freedom” – in Lecture IX of *Political Liberalism*: “Political autonomy is specified in terms of various political institutions and practices, as well as expressed in certain political virtues of citizens in their thought and conduct – their discussions, deliberations and decisions – in carrying out a constitutional regime” (*PL*, §IX.3.3). Isn’t this image of the requirements of committed citizenship (i.e. “political autonomy”) – or, put differently, the demands of robust reasonableness – a comprehensive (not merely political) image of human flourishing? Below, we will address this very important question in detail.

Despite such concerns, this chapter, like the earlier chapters, is ultimately a *defence* of the Rawlsian political enterprise. What we shall ultimately find is that political liberalism is *not* neutral among comprehensive doctrines. *But that is not its aim:* political liberalism need only be neutral among *reasonable* comprehensive doctrines, and in this I believe that Rawls succeeds. The distinction between reasonable and unreasonable doctrines is the subject of a lengthy discussion below. It is also the case that Rawls *does* rely on the prevalence of fuller, comprehensive conceptions of the good but he does so on a conscientious, principled basis (as the most likely route to stability for the right reasons). Again: the possibility of political liberalism explicitly depends on the widespread commitment to citizenship becoming embedded in the various comprehensive doctrines. In other words, a “thicker” image of human flourishing *is* a part of this doctrine, but it is not “smuggled in” from the outside: like justice as fairness before it, political liberalism looks backwards, trying to articulate an overlapping consensus of political principles to which the holders of various comprehensive doctrines can already give their assent *based upon* – or, put differently, from within – their deep (and deeply held) existential commitments. Below, we will try to get a better understanding of what it means for the political conception of justice to be “freestanding,” and to show that Rawls’s terminology here is a bit misleading (but is ultimately *not* incoherent or a violation of the fundamental principle at the heart of political liberalism).

And so, just as earlier chapters were devoted to the development of a more robust conception of the requirements of reasonableness, this final chapter will attempt to develop a more robust, and therefore convincing, conception of the idea of public reason. Hopefully, this should assuage
some of the critical concerns over the neutrality (and, by extension, coherence) of Rawls’s political philosophy. What we shall ultimately find is that the Rousseauian and Hegelian heritage of Rawls’s normative philosophy leads us to two important conclusions: (1) that political liberalism, as described by Rawls, is in need of significant expansion (or elaboration), but (2) that those revisions ought to be thought of as *internal* to the idea of public reason at the centre of that doctrine. In the end, and in light of the conception of robust reasonableness developed in earlier chapters, it is clear that citizens in the kind of political community described by Rawls *can* handle deep, irreconcilable differences without resorting to either institutionalized processes of exclusion or violence (however remote the latter possibility). The ultimate purpose of this final chapter, then, is to show that Rawls himself has very compelling reasons to endorse the revised image of political deliberation outlined below.

I.

Does Rawls’s (eventual) distinction between political and comprehensive doctrines provide him with the necessary resources to respond to the charge that his claims to neutrality ultimately fail? In *Political Liberalism*, Rawls claims that, despite his (sustained) emphasis on the priority of right over the good, his own political doctrine does make reference to five essential (though ultimately neutral) ideas of the good.\(^{92}\) They are listed in Lecture V of *Political Liberalism*. They are: (1) the idea of goodness as rationality, or the supposition that “the members of a democratic society have a rational plan of life in the light of which they schedule their more important endeavours and allocate their resources so as to pursue their conceptions of the good over a complete life” (*PL*, V.2.1) (2) the idea of primary goods, which supplies “a practicable public basis of interpersonal comparisons based on objective features of citizens’ social circumstances” (ibid., V.3.4) (3) a distinction between permissible and impermissible conceptions of the good, which is premised on the compatibility of a particular doctrine with a respect for the rights of all citizens (i.e. with mutual toleration) (ibid., V.5) (4) a conception of

\(^{92}\) Only the first two of these (supposedly neutral) conceptions of the good were present in *Theory*, which confirmed the suspicions of those who believed that the earlier text presupposed a fuller normative ideal than Rawls’s emphasis on a “*thin*” theory of the good claimed. See Mulhall and Swift 2003, p. 481.
the political virtues “so that [citizens] want to honour the fair terms of social cooperation in their relations with the rest of society” (ibid., V. 6.3) and (5) the good of political society: “the good that citizens realize both as persons and as a corporate body in maintaining a just constitutional regime and in conducting their affairs” (ibid., V.7.1). Indeed, as we have already seen, political liberalism – like justice as fairness before it – endorses an image of the political community where citizens do have common (and not merely instrumental) ends: the maintenance of just institutions and adhering to the dictates of robust reasonableness (as described in Chapter 4).

According to Rawls, these ideas of the good are compatible with the kind of neutrality Political Liberalism seeks to defend: they are purely political values – they apply only to the sphere of politics and are based upon ideas and values present in the public political culture – and are therefore compatible with a multiplicity of reasonable comprehensive doctrines. In other words, the affirmation of the five goods of political liberalism does not presuppose (and thereby validate as true) a particular comprehensive doctrine. Instead, it creates a common ground – an overlapping consensus – between the various reasonable comprehensive conceptions held by (and that could be held by) citizens (ibid., §IV.3). Such is the only reasonable response to the inevitable (reasonable) pluralism of the modern milieu; any doctrine that fails to respect the inevitable fact of pluralism is itself a failure as a guide for political life.

In this light, it is easy to see how the earlier liberalism of Theory was itself comprehensive and therefore sectarian in light of its institutionalized status: Theory, that is, sought to license the exercise of the coercive power of the state in ways that cannot be justified (as reasonable) to all citizens (i.e. according to the tenets of justice as fairness, which privileges the value of rational, or Kantian, autonomy) (Freeman 2003, p. 30 – 35). Liberalism, like any religious worldview, is a sectarian doctrine, and so is equally a threat to the stability of the plural community. “This,”

---

93 It must be admitted that, in Political Liberalism, determining what constitutes “the political” is an ambiguous matter. According to Rawls, the values of political liberalism apply solely to “constitutional essentials and questions of basic justice;” and so, by extension, those issues which do not fall under either heading need not be settled solely with reference to the purely political values of political liberalism: “Many if not most political questions do not concern those fundamental matters, for example, much tax legislation and many laws regulating property; statutes protecting the environment and controlling pollution; establishing national parks and preserving wilderness areas and animal and plant species; and laying aside funds for museums and the arts” (PL, VI.5). But it remains unclear how exactly to draw the between “fundamental matters of justice” and the rest (see e.g. Greenawalt 1988, p. 226 – 227, Greenawalt 1994, p. 682 – 688 and Beiner 2011, Ch. 23).
says Beiner, “is unquestionably the core idea of Political Liberalism” (Beiner 2011, p. 285; see also Habermas 1998, p. 86). And so, comprehensive liberals (such as Kant and Mill), for whom autonomy is the ultimate – the true – expression of human nature, are asked to show (discursive) restraint when dealing with their fellows on matter of public (constitutional) justice.

But this distinction between political and comprehensive doctrines leads to an obvious, and important, question: how can Rawls rebut those who reject the values of political liberalism – or, in another obvious and likely possibility, those who accord the values of political liberalism less weight than the values of their (potentially illiberal) comprehensive doctrine, and who therefore wish to see this parochial set of values institutionalized in the basic structure – without dismissing those (comprehensive) doctrines as false. After all, political liberalism itself (as a political doctrine) is an expression of Rawls’s ultimate philosophical agnosticism on such comprehensive matters: it is precisely because we lack the intellectual resources to verify the essential truth of this or that doctrine that we confine philosophy to the comparatively “shallow” sphere of politics (see PL, II.2.1; see also Raz 1990, p. 8 – 9).

Rawls’s response to these insuperable epistemological limitations is to dismiss those comprehensive doctrines with political ambitions not as false, but as unreasonable: unreasonable doctrine holders seek to subject their fellows to the coercive exercise of state power according to sectarian values and principles (i.e. according to principles to which only some can consent); these citizens fail to recognize the burdens of judgment, or the inevitable existence of the sources of reasonable disagreement (among reasonable persons) about the appropriate content of a comprehensive doctrine (ibid.; see also Greenawalt 1994, p. 670 – 679). Recall, in this context, the Kantian rendition of the demands of reasonableness: “Persons are reasonable in one basic aspect when, among equals say, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so […] Reasonable persons desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept” (ibid., II.1.1). This is the liberal principle of legitimacy (ibid., §VI.2.1).
Reasonableness, in other words, is agreement with the values and principles of political liberalism\(^94\) – with the belief that coercive power ought to be yielded in certain (universally justifiable) ways and that, in turn, citizens’ obligations towards their fellows are characterized by the existence of certain constraints on action and on reasons (Mulhall and Swift 2003, p. 482; see also Greenawalt 1994, p. 671 – 672 and Bohman and Richardson 2009, p. 259 – 260 and p. 266). Rawls thus rules out the possibility of reasonable disagreement over political principles: reasonableness is agreement with the prevailing public cultural conception of social cooperation and of the demands of citizenship (i.e. with political liberalism) (\(PL\), §I.2.2; see also Habermas 1998, p. 83). “There is no such thing,” adds Waldron, “as reasonable disagreement in politics” (Waldron 1994, p. 376; see also Habermas 1998, p. 86 – 89, Scheffler 1994, p. 20, McCarthy 1994, p. 51, Cohen 2009, p. 30 and Beiner 2011, p. 297). The only instance of reasonable disagreement at the level of politics, by extension, is disagreement about what the application of the principles of justice requires.\(^95\)

This raises an obvious question: how exactly can Rawls defend the primacy of his political conception? It is not immediately obvious why the values of political liberalism necessarily trump those other, comprehensive sets of values: what extra work does their status as public (i.e. institutionalized) principles do? Why, in other words, would comprehensive doctrine holders privilege the needs of citizenship over the demands of faith? Says Beiner: “It starts to look as if […] key aspects of [comprehensive] liberalism [i.e. the view that citizenship is sufficiently important that religious commitments, and other existential commitments, should not trump a commitment to the kind of ecumenical citizenship characteristic of justice as fairness] have been slipped in the back door” (Beiner 2010, p. 286; see also Hollenbach 1993, p. 889, Scheffler 1994, p. 13 – 14 and p. 16 and McCarthy 1994, p. 51 – 54). According to this line of criticism, political liberalism ultimately relies on those comprehensively liberal doctrines it was explicitly

---

\(^94\) Rawls acknowledges the likelihood of controversy over the use of the principles of justice as fairness as the standard for public reason: “It is inevitable and often desirable that citizens have different views as to the most appropriate political conception; for the public political culture is bound to contain different fundamental ideas that can be developed in different ways” (\(PL\), §VI.4.5 italics added).

\(^95\) As we shall see below, in Section III, this kind of disagreement (about what the application of justice principles requires) can be quite deep. See below.
designed to exclude from public political debate (in light of its constitutive concern with neutrality).

Indeed, in Lecture IV of *Political Liberalism*, Rawls makes an apparent concession along these lines. In a discussion of the unreasonable comprehensive doctrine holder – someone who insists on the *truth* of their doctrine, and on its appropriateness as a guide for civic life, and who actively attempts to convert others, whether through violence or through institutional socialization – Rawls offers the following advice: “We say of this [sectarian doctrine] that [those who hold it] are mistaken [in their denial of the existence of reasonable pluralism]; but we need not say that their religious beliefs are not true” (*PL*, IV.3.3). So far, this is reconcilable with Rawls’s discussion of the burdens of judgment and with the political character – or, the agnostic attitude – of political liberalism. But Rawls continues: “Of course, we do not believe the doctrine believers here assert, and this is shown in what we do. Even if we do not, say, hold some form of the doctrine of free religious faith that supports equal liberty of conscience, our actions nevertheless imply that we believe the concern for salvation does not require anything incompatible with liberty. Still, *we do not put forward more of our comprehensive view than we think needed or useful for the political aim of consensus* [… we may be forced to refer, at least in some way, to such a [comprehensive] view” (ibid. and §IV.5.1 italics added; see also Haldane 1996, p. 66 and Cohen 2009, p. 31).

This is an important and revealing passage. In it – specifically, in the italicized section – Rawls acknowledges the lack of any non-comprehensive resources for dealing with the sectarian denial of pluralism; and so, by extension, it seems to be permissible to *publicly* defend the political conception of justice with reference to one’s comprehensive (but importantly liberal-friendly) doctrine. This is going to be a crucial aspect of the expanded doctrine of public reason that I present below in Sections II and III. But, for now, it is important to take note of the fact that political liberalism is starting to look like an *element* of one’s comprehensive doctrine – whether liberal, religious or secular – as opposed to a “freestanding” doctrine characterized by its inattention or indifference towards – its transcendence of – competing conceptions of the good.

In fact, I hope to show that, according to Rawls, the fact (and the significance) of reasonable pluralism cannot be recognized independently of one’s comprehensive commitments. Consider this important passage from *Political Liberalism*: “In a modern society with its numerous offices
and positions, its various divisions of labour, its many social groups and their ethnic variety, citizens’ total experiences are disparate enough for their judgments to diverge on many if not most cases of any significant complexity” (*PL*, II.2.3, which contains a detailed list of the deeply personal, idiosyncratic ways in which individuals experience, and weigh the evidence of, the world). Rawls thus recognizes that a partial or parochial view of the world is an insuperable psychological fact: one is constantly confronted by an ideationally (comprehensively) hostile world (i.e. a world characterized by comprehensive pluralism). There is, by extension, no sense in which one must (or even could) transcend (or bracket or ignore) one’s comprehensive commitments when engaging in political activity or public debate. In other words, adds Hampton, the possibility of a stable overlapping consensus precisely depends on the contingent fact that “each party has a metaphysics that supports the ideas in the shared fund as ‘right’” (Hampton 1989, p. 805; see also Habermas 1995, p. 122 – 126).

This is a very important point: when it comes to the daily practice of public reason by engaged individuals, there is, on Rawls’s own terms, no value-free (or “freestanding”) space to which individuals can (or must) “ascend” before entering the public sphere. Again, and as we shall see in detail below, recognizing the dictates of reasonableness is a by-product of the contingent evolution of one’s comprehensive doctrine. Here is a typical statement of this view from “The Idea of Public Reason Revisited”: “[T]he roots of democratic citizens’ allegiance to their political conception lie in their respective comprehensive doctrines, both religious and nonreligious. In this way citizens’ allegiance to the democratic ideal of public reason is strengthened for the right reasons. We may think of the reasonable comprehensive doctrines that support society’s reasonable political conceptions as those conceptions’ vital social basis, giving them enduring strength and vigor” (Rawls 1999c, p. 153). In other words, a shared political conception of justice is affirmed from within each of society’s comprehensive doctrines.

Habermas, like Hampton, is therefore exactly right when he claims that, in political liberalism, “the metaphysical […] remains the ultimate ground of the validity of what is morally right [i.e. the principles of justice]” (Habermas 1998, p. 85; see also ibid., p. 76). There is no disagreement with Rawls on this question. But, of course, Habermas intends for this passage to be a critique: his ultimate concern is that, by (apparently) excluding comprehensive reasons from the public sphere, political liberalism produces the “counterintuitive” result that “a public conception of
justice [...] ultimately derive[s] its moral authority from non-public reasons. Anything valid should also be capable of public justification” (ibid., p. 86).

Critics like Habermas are troubled by what appears to be Rawls’s prima facie exclusion from public debate of those with strong comprehensive convictions (see e.g. Greenawalt 1994, p. 678 – 684, Sandel 1994, p. 1768 – 1772 and Habermas 1995, p. 126 – 130). The problem is compounded by the fact that Rawls clearly relies on these comprehensive convictions for support of the political conception of justice. Indeed, this is perhaps the most frequent critique of Political Liberalism: that Rawls either excludes comprehensive doctrines from the public sphere altogether or (equally problematically) that the adherents of comprehensive doctrines must speak the non-sectarian language of political liberalism in public, thus compromising their connection to what are existentially deeper values and commitments (and also compromising the likelihood of stability for the right reasons). Of course, all of this raises an obvious prior question: exactly how restrictive is Rawls’s position on the use of (say) religious reasons in public political debate? This question is the subject of Section II.

What we shall ultimately find is this: that political liberalism is dependent for support – for stability for the right reasons – on liberal-friendly comprehensive doctrines but this is not synonymous with its being dependent on the universal adoption (and subterranean inclusion) of comprehensive liberalism. As we shall see, comprehensive liberalism is not “slipped in the back door;” rather, reasonableness is a constitutive element of various comprehensive doctrines, and abiding by its constraints is a meaningful expression of one’s commitment to the dictates of one’s comprehensive doctrine. This is precisely why (contra Habermas) Rawls is willing to admit comprehensive concerns and reasons into public political debate. Ultimately, the main aim of Sections II and III96 is to show the compatibility of Rawls’s theory with a maximally inclusive, maximally permissive public sphere.

96 Both of these sections have benefited immensely from discussion with Bruce Lyth and from his paper “The Civil Religion of John Rawls.”
II.

According to Gaus and Vallier, Rawls’s position on public reason is in fact highly “permissive:” “[Rawls] advocates the ‘wide view’ of public reason, which means, more or less, that citizens can rely on religious reasons to motivate or justify support of particular policies so long as adequate public [i.e. non-religious, universally endorsable] justification is forthcoming” (Gaus and Vallier 2009, p. 56). This is the Rawlsian proviso, or the “wide view” of the public political culture: “Reasonable comprehensive doctrines, religious or non-religious,” he says, “may be introduced in public political discussion at any time, provided that in due course proper political reasons – and not reasons given solely by comprehensive doctrines – are presented that are sufficient to support what the comprehensive doctrines are said to support” (Rawls 1999c, §4; see also Hertzberg forthcoming, Ch. 3, §III).

For Rawls’s religious critics, the necessity of eventual translation into the public language of political principles is highly problematic: “It belongs to religious convictions of a good many religious people in our society that they ought to base their decisions concerning fundamental issues of justice on their religious convictions. They do not view it as an option whether or not to do so. It is their conviction that they ought to strive for wholeness, integrity, integration in their lives” (Audi and Wolterstorff 1997, p. 105; see also Habermas 1998b, p. 126 – 130 and Rorty 2003, p. 142 – 144). In other words, the truth of our comprehensive convictions ought to have a place in the public sphere as the whole basis of our adherence to, or rejection of, a given law or policy; after all, adds Josh Cohen, “it is hard to understand what it means to drop the concept [of truth] (unless we are also prepared to drop, among other things, the concept of belief, assertion, judgment)” (Cohen 1997, p. 41). Religious citizens regard revelation as the appropriate basis of social and political legislation given the status of their religious beliefs as true. And so, in the interest of equal respect, it seems to be the case that we ought not prohibit altogether proclamations of truth from the public sphere.\(^\text{97}\)

Here, we return to our earlier brief mention of the unequal cognitive burden placed on religious citizens: while they may now (with the later addition of the proviso) be able to speak religion in

---

\(^{97}\) As we shall see below, the theoretical challenge is to identify the proper limits of the justificatory power of religious beliefs based on revelation.
the political public sphere – that is, to make reference to the religious grounding of their preferred laws and policies – they are still under an obligation (of civility) to (eventually) “translate” those reasons into the familiar language of the political public culture and of the justice principles that characterizes that culture. In other words, many of Rawls’s critics regard this move as the extension of a superficial kind of respect to religious believers: it remains the case that religious claims cannot inform policy all the way down, and for many faiths this is often an irreconcilable demand – one’s faith, that is, ought to be the exclusive ground for one’s approval of, and adherence to, a given law or policy. Even with the later addition of the proviso, for certain citizens there remain burdensome hurdles for entry into public political debate.

It is my view that Rawls’s critics’ emphasis on the problems associated with translation misses the importance of the wide view of public reason (as outlined in “The Idea of Public Reason Revisited”). Simply put: the proviso is not about the necessity of eventual translation into the language of the political public culture. In fact, the notion of translation fundamentally misses the point: the usefulness of the proviso does not depend on the ability of individuals to translate their religious convictions into public principles. The proviso, rather, is much less ambitious and much more permissive (and therefore much less controversial): the wide view of public reason does not even require ordinary citizens to speak the language of the political conception of justice. Says Rawls: “The idea of public reason does not apply to the background culture with its many forms of non-public reason. Sometimes those who appear to reject public reason actually mean to assert the need for full and open discussion in the background culture. With this political liberalism fully agrees” (Rawls 1999c, p. 134; see also PL, §1.2.3).

98 In a forthcoming article, Walhof raises important and interesting questions regarding this notion of “translation”: “Translation implies an equivalence of some sort. When we attempt to translate a statement from one language into another, for example, the task is oriented toward conveying the same meaning in the second language, even though there is, of course, an interpretive dimension to any translation. But [the Rawlsian argument for the proviso, and other arguments that emphasize the translatable nature of public claims, does not] really [describe] religious arguments restated in nonreligious terms. There is not an equivalent meaning simply expressed in two different ways. Instead, the given arguments are facially nonreligious reasons that are situated within broader theological-cum-political visions. In fact, the stated reasons make the most sense, and are most compelling, when one also knows these broader visions. Moreover, such contextualized arguments would in many respects be more accessible to one who does not share these visions than the stripped down versions offered” (Walhof forthcoming). See also Chambers 2007, p. 214: “The problem with [the translation proviso] is not simply that such scientific/constitutional language falsifies the beliefs of the speaker, that it places an unfair translation burden on the speaker, or even that it is clearly a pseudo-language adopted for strategic reasons and so subject to questions of sincerity. The problem is that meaning is lost.”
This distinction between the background culture, which Rawls elsewhere describes as “the culture of civil society,” and the political culture is crucial to a proper understanding of the wide view of public reason. The constraints of public reason only apply to the latter: when dealing with matters of basic constitutional essentials – with the design, implementation and reform of the basic institutional structure of political society – judges, legislators and chief executives must proceed entirely within a political conception of justice characterized by conspicuously political values. Such values might include, for example, “a more perfect union, justice, domestic tranquillity, the common defense [sic], the general welfare, and the blessings of liberty for ourselves and our posterity” (Rawls 1999c, p. 144). By contrast, in the background culture of civil society – where the members of “churches and universities, learned and scientific societies, and clubs and teams” debate in what Rawls refers to as “the public square” or “the public forum” – citizens are free – in fact, they are encouraged by Rawls – to speak the (sectarian) language of their particular comprehensive worldview (PL, §I.2.3 and Rawls 1999c, p. 134). There is an important distinction, in other words, between “a religious leader’s preaching or promoting a pro-life organization [which is permitted according to this wide view] and leading a major political movement or running for political office” (Greenawalt 1988, p. 226). In the end, religious reasons are intelligible to secular citizens (as reasons), even though they may not be (at bottom) rationally or discursively redeemable (Rawls 1999c, p. 153). Simply put: recourse to religious reasons will not shock or offend modern liberal citizens.

99 In other words, the proviso only applies at a later, institutional stage: it does not (to reiterate) apply to the public pronouncements of comprehensive doctrine holders in the public sphere; hence, its characterization by Gaus and Vallier as “permissive.” This raises a series of important questions: is this, in the end, a revision or expansion of the wide view of public reason? Or, is such a view of the proviso implicit in “The Idea of Public Reason Revisited”? This is an ambiguous matter, because Rawls himself does not provide much guidance regarding when – at what stage – the proviso ought to be applied: “Obviously, many question may be raised about how to satisfy the proviso. One is: when does it need to be satisfied? On the same day or some later day? Also, on whom does the obligation to honor [sic] it fall? It is important that it be clear and established that the proviso is to be appropriately satisfied in good faith. Yet the details about how to satisfy it must be worked out in practice and cannot feasibly be governed by a clear family of rules given in advance” (Rawls 1999c, p. 153). Again, it is my view that the proviso ought to be applied at a later, institutional stage by the elected and appointed actors that represent the basic structure of the political community. The by-product of such a view is a maximally inclusive, maximally permissive political public sphere: this is precisely to efface critical concerns about the troubles associated with the “translation” of comprehensive commitments into the language of public reason. We consider the implications of my own view – which, again, can be conceptualized as either a revision or an elaboration (depending on one’s own view of what Rawls means by “in good faith”) – below in Section III.
In fact, in most cases the political conception is “supported by a reasonable overlapping consensus of such doctrines” (Rawls 1999c, p. 143). Indeed, for Rawls, those cases of radical incompatibility between comprehensive convictions and public reason are rare: this is perhaps the ultimate testament of the capacity of liberal democratic principles to modernize those comprehensive beliefs that were previously characterized by the desire to “win the world for the whole truth” (Rawls 1999c, p. 132). Weithman expresses this idea nicely: “In an overlapping consensus, adherents of diverse philosophical and religious traditions accept justice as fairness [or some other political conception], but for different moral reasons; each supports the conception for reasons drawn from his or her own more comprehensive moral views. Thus, in an overlapping consensus, Kantians affirm justice as fairness for one set of reasons, utilitarians accept it for another, and Christians accept it for still another.” (Weithman 1994, p. 197; see also ibid., p. 199, Greenawalt 1994, p. 679 and Hollenbach 1993, p. 882 – 889 esp. p. 887). Acclimatizing believers to the fact of pluralism is, to reiterate, liberalism’s greatest historical accomplishment; this is the tumultuous history of which Political Liberalism is the attempted systemization.

This is a very important point: political liberalism is best understood as a response to – and an attempt to take full advantage of – the contingent but ultimately favourable evolution of a plural society’s reasonable comprehensive doctrines; that is, the primacy of ecumenical citizenship (over revelation, say) is not a condition that is imposed from the outside by the political philosopher (as a precondition of political participation). Rather, the willingness to privilege reasonable citizenship is a constitutive dimension of the reasonable comprehensive doctrine at the centre of one’s existential universe. In other words, it is not a question of choosing between politics and religion: reasonable (i.e. tolerant) political activity is precisely an expression of one’s commitment to one’s comprehensive beliefs, not the “bracketing” out of the latter when one enters the public sphere. Indeed, in Political Liberalism, there is no sense in which one is

100 According to Rawls, it is the responsibility of judges, legislators and chief executives to ensure that citizens come to regard their religious beliefs as compatible with public principle, with the language and values of public reason (Rawls 1999c, p. 165).

101 Even though the following passage is intended as a critique of Rawls, it is becoming clear that there is no fundamental disagreement between Rawls and Waldron on this question of the political direction (or implications) of a comprehensive doctrine. See Waldron 1994, p. 383: “If a religious or philosophical tradition has nurtured a rich
forced to choose between the public and private spheres (or selves): the precise nature of the relationship between the two is determined by the organic evolution of a given comprehensive doctrine as it grapples with the fact of deep pluralism. And so, like Raz, Rawls recognizes that “there can be no justice without truth” (Raz 1990, p. 15; see also Hampton 1989, p. 807 and Habermas 1998, p. 87). But Rawls gives this belief a unique, practical rendition: that is, we need not think of the principles of justice that animate our political conception of justice as true. Rather, their compatibility with – no, the fact that they are the organic outgrowth of – the truth of our comprehensive commitments makes them the most suitable basis for political stability.

Of course, we do not even need to go that far (i.e. towards deep or even complete compatibility): a clear instantiation of civic virtue is the recognition of the deep incompatibility of the tenets of one’s comprehensive doctrine with this or that public policy with the simultaneous acceptance of said policy as the most reasonable option given the radical pluralism of the political community (Rawls 1999c, p. 137). The Catholic citizen, for example, can continue to voice his or her opposition to abortion while recognizing the legitimacy of the liberal state that permits this procedure (as the most reasonable option given the dominant ordering of society’s political values). Devout Catholicism (as an exemplar of a thoroughly systematized religious worldview) can co-exist with liberal citizenship even in those instances where these two commitments lead to opposite policy prescriptions (ibid., p. 170). This is, once again, the major psychological accomplishment that makes radical pluralism (politically) manageable.

The example of Catholic opposition to abortion raises a number of difficult questions: Is it realistic to advocate a form of politics that enforces a deep discontinuity between people’s self-understanding in private and political life? Is this kind of existential distancing even psychologically possible? Ultimately, I believe that this is an empirical question: it may or may not be the case that (potentially very deep) discord between one’s public and private beliefs will

and resourceful conception of the good, it would be odd to expect its priests, ideologues, or philosophers not to have developed that conception also in a social or political direction. Social and political concerns after all are among the most pressing concerns we have: it would be odd if a tradition had views about what made life worth living but no views about the basis on which we ought to live together [...] ‘Comprehensive’ seems to indicate an ambition on the part of such conceptions to answer all the big questions, from which questions of justice and the basic organization of society can hardly be excluded.”
lead to a kind of existential or psychological schizophrenia. My suspicion, though, is that one’s commitments as citizen and as believer are relatively easily managed, and this is especially the case in the liberal milieu where freedoms of conviction and association are institutionalized in the basic structure. Indeed, the history of plural liberal democracies has provided a compelling body of evidence that suggests that the management of these psychological complexities is hardly an insuperable problem (see e.g. Macedo 1998, p. 61 and p. 72 – 73).

However, in the event that one’s comprehensive commitments are fully irreconcilable with the public political culture – again, this is regarded by Rawls as a highly remote possibility in light of his own recurring emphasis on the socializing tendencies of liberal principles of justice – one is free to recede from public life. Here, Rawlsian public reason emerges as a fairly muscular doctrine: it has no patience with – it excludes on principle – the unreasonable and the intolerant (Rawls 1999c, p. 138; see also Macedo 1995, p. 470 and Macedo 2000, Pt. II). And so, those obdurate comprehensive doctrines that are immune to processes of modernization – which is to say, all those comprehensive systems of belief that are doctrinally hostile to the values of freedom, equality, universal self-respect and toleration that animate political liberalism – have no place in the Rawlsian discursive milieu. There is simply no getting around this; here, we confront the principled limits of Rawlsian neutrality. But, again, these recalcitrant doctrines are regarded by Rawls as the exception: there is not an unbridgeable chasm, in most cases, between one’s comprehensive beliefs and the kinds of foundational principles that make for good (liberal) citizenship. This is the most important lesson of “The Idea of Public Reason Revisited,” with its emphasis on the wide view of public reason.

III.

How, then, should we proceed? As we have just seen, the appeal to religious reasons – to revelation – is acceptable in the Rawlsian public sphere. Indeed, Rawlsian public reason is (for many) unexpectedly permissive: in civil society, ordinary citizens are not required to translate their comprehensive reasons into the political language of public reason. Instead, in Rawls’s scheme, the obligation to fulfil the proviso falls primarily on elected and appointed institutional actors – the representatives of the basic structure, and of the ordering of political values that characterizes that structure, such as chief executives, elected legislators and judges: “[These
institutional actors] argue from what [they] believe, or conjecture, are other people’s basic doctrines, religious or secular, and try to show that, despite what they may think, they can still endorse a reasonable political conception that provide a basis for public reason” (Rawls 1999c, p. 156). After all, in order to regard themselves as the proper subject of state coercion, religious citizens – like all citizens – must come to see their values embodies in the laws and policies of the state. This is (recall) the liberal principle of state legitimacy, and it is the institutionalized responsibility of state actors to bring about an ideational condition characterized by the stable, principled compatibility of comprehensive and political values.

The state of affairs to which Political Liberalism aspires, then, is one in which the plural comprehensive commitments of citizens are deeply, thoroughly compatible with the guiding values and principles of justice as fairness. This is, to reiterate, the desirable outcome of the history of liberal democracy into which Rawls inserts his own thought in the Introduction to Political Liberalism. But there are two important caveats that Rawls includes within this wide view of public reason: (1) just as the representatives of state institutions are under an obligation (of reasonableness) to give appropriate hearing to the comprehensive interests and reasons of religious citizens, so too are comprehensive doctrine holders under a moral (not legal) obligation of civility to give public reasons for their political policy preferences (Rawls 1999c, p. 135 – 137). And (2) there are important limits to the comprehensive genesis or heritage of state principles and actions. Let us deal with the second condition first.

According to Rawls, religious (and other comprehensive) obligations that involve the violation of the rights and liberties of other citizens have no place in the political public sphere (ibid., p. 138). In other words, religious reasons must be compatible with the first principle of justice – they must be reasonable, they must treat all persons as free and equal, they must accept the fact of pluralism and tolerate the so-called apostasy of those with different systems of belief (Hertzberg forthcoming, Ch. 3, §I; see also Cooke 2007, p. 233 – 235, Spragens 2008, p. 207 and Cohen 2009, p. 33 – 35). And, of course, Rawls expects that this will be the case in the liberal democratic milieu: well-ordered institutions tend to liberalize the authoritarian and persecuting tendencies of the comprehensive doctrines with which they come into contact. Even without this emphasis on the socializing capacity of liberal principles and institutions, adds Cooke, “there is no conflict in principle between non-authoritarian reasoning and an orientation towards some ‘otherworldly,’ transcendental source of validity” (Cooke 2007, p. 235). Religion can foster
good citizenship. Political liberalism, as we have just seen, depends on this possibility *on principle*. However, given the prevailing, overwhelmingly secular character of most modern liberal democracies, those laws and policies for which there are *only* religious reasons are illegitimate.\(^{102}\) We must be careful not merely to invert the problem associated with the authorship requirement of all legitimate law: secular citizens must not come to see themselves at the mercy of an alien coercive authority (that takes its political cue from sectarian revelation).

And so, according to Cooke, “a religious citizen must always believe that there is a reasonable secular rationale for any law he or she proposes, even though his or her grounds are thoroughly religious” (ibid., p. 62; see also Rawls 1999c, §6.1). Indeed, it is likely the case that such an accessible, public rationale exists, for “even citizens who reason on religious grounds share secular concerns: health, housing, earning and protecting income and public safety – laws that appeal to these are often endorsed by all members of the public” (ibid., p. 61; see also Rawls 1999c, p. 173). Adds Hertzberg, drawing an illuminating contrast between agenda setting and policy and legal resolution: “The conscientious objection to a [liberal] mode of inquiry, in the end is reasonable – so long as it arises from a desire to *participate* in politics on religious grounds alone and not a desire to *control* politics on religious grounds alone” (Hertzberg forthcoming, Ch. 3, §V).

Regarding the second caveat – the moral obligation of civility to use the values and language of public reason, regardless of one’s religious, secular or political orientation – the image of society

---

\(^{102}\) One very promising alternative can be found in Hertzberg’s forthcoming dissertation (especially Ch. 3, §II): there, Hertzberg argues that Rawlsian public reason norms ought not apply to the agenda setting phase of public deliberation (i.e. those deliberative engagements where citizens select the issues and proposals they will discuss together). What Hertzberg attempts to articulate, then, is a new (less restrictive) substantive norm to guide the earlier, agenda setting phase of public debate; and what he ends up with is quite similar to the Habermasian view of (nearly) uninhibited public deliberation described above: citizens who argue that a given issue or policy should be put on the public agenda *can* appeal to their reasonable comprehensive doctrine (without any accompanying processes of translation into the language of political principles). Of course, the caveat of reasonableness is designed to exclude those authoritarian doctrines that may de-rail productive deliberation and undermine the sense of community; but, as we have also just seen, the majority of comprehensive (religious) doctrines in the well-ordered liberal society are reasonable (i.e. reconcilable with an extensive scheme of equal basic liberties). Ultimately, though, the resolution of these public debates must be capable of articulation in the universally accessible language of public political principles: “[deliberators] only pass resolutions that can be justified by appeal to a reasonable political conception of justice […] citizens must recognize the force of public reason as a resolution norm.” According to Hertzberg, this emphasis on the *phases* of public reasoning – (comprehensively permissive) agenda setting and (more restrictive processes of) resolution – satisfies the essential criteria of Rawlsian public reason.
as a cooperative venture is nicely captured by this notion of mutually burdensome public deliberation. This is, as per the argument of Chapter 2, part of the Hegelian heritage of justice as fairness. Similarly, the wide view of public reason is, in my view, a fuller expression of the ethos of mutual respect – and of equal mutual recognition – that informs our shared institutions and upon which our sense of community rests. This is, as per the argument of Chapters 3 and 4, part of the Rousseauian heritage of justice as fairness. Indeed, the inclusion of religious citizens in the public sphere – without the caveat of obligatory translation into the language of justice as fairness (or some other liberal doctrine) – honours both “the value of tolerance” and (perhaps more importantly) “the ideal of the political arena as a space of public reasoning open to equal citizens” (Cohen 2009, p. 7).

Like Rawls, Bohman and Richardson endorse a similarly wide view of public reason with its emphasis on the maximally inclusive, maximally permissive public sphere: “Deliberation requires citizens to engage with one another’s view civilly, forthrightly addressing them rather than passing over them in silence, and engaging them rather than treating them as anthropological curiosities […] one may attempt to enter into the other’s perspective, not so as to critique it, but more open-endedly to attempt to understand it” (Bohman and Richardson 2009, p. 271 – 272; see also Tully 2005, ch. 4). 103 Habermas too emphasizes the mutually burdensome character of public deliberation in the comprehensively plural milieu: “The recognition by secular citizens that they live in a post-secular society that is also epistemically attuned to the continued existence of religious communities is a consequence of a change in mentality that is no less cognitively exacting than the adaptation of religious consciousness to the challenges of an environment that is becoming progressively more secular […] secular citizens are willing to enter into a political discussion of the content of religious contributions with the intention of translating potentially morally convincing intuitions and reasons into generally accessible language.” (Habermas 1998b, p. 139).

103 See also Tully 1995, p. 17 – 29 and Ch. 4. Says Tully in a representative passage: “If there is to be a post-imperial dialogue on the just constitution of culturally diverse societies, the dialogue must be one in which the participants are recognised and speak in their own languages and customary ways. They do not wish to be silenced or to be recognised and constrained to speak within the institutions and traditions of interpretation of the imperial constitutions that have been imposed over them. This world reversal, from a habitual imperial stance, where one’s own customary forms of reflection set the terms of discussion, to a genuinely inter-cultural popular sovereignty, where each listens to the voices of the others in their own terms, is the most important and difficult step in contemporary constitutionalism.”
For deliberative democrats, this image of inclusive deliberation aimed at mutual understanding – of the maximally inclusive public sphere – is itself the basis for solidarity in a plural milieu. Says Habermas: “If we understand the democratic procedure as a method by which legitimacy is generated from legality, […] then the issue of a deficit of validity [which was previously premised on the “pre-political ethical convictions of religious or national communities’’] never arises […] constitutional patriotic bonds form and renew themselves in the medium of politics itself” (Habermas 1998a, p. 104 and p. 106; see also Habermas 1998b, p. 120 – 122). And Cooke draws the important connection between equal public recognition and political legitimacy: “The particular cost of […] [the] requirement of translation into secular language is damage to the principle of equal respect for the political autonomy of citizens” (Cooke 2007, p. 231). None of this, it seems to me, is irreconcilable with Rawls’s own understanding of the nature of deliberative democracy. Recall the necessary connection drawn by Rawls between institutional recognition and self-respect: if one cannot see oneself as the co-author of the laws of one’s political community, one’s self-respect is damaged, and perhaps irreparably so (Rawls 1999c, p. 137).

Indeed, in light of the wide, permissive view of public reason outlined here, it is clear that the distance between Rawls and his deliberative democratic critics is significantly reduced. Like Habermas and Cooke, Rawls also regards inclusive discursive participation as a crucial source of social stability and vitality: “When firm and widespread, the disposition of citizens to view themselves as ideal legislators, and to repudiate government officials and candidates for public office who violate public reason, is one of the political and social roots of democracy, and is vital to its enduring strength and vigor” (Rawls 1999c, p. 135; see also ibid., pp. 170 – 171 and p. 175). Again, Rawls: “When citizens deliberate, they exchange views and debate their supporting reasons concerning public political questions. They suppose that their public opinions may be revised by discussions with other citizens; and therefore these opinions are not simply a fixed outcome of their existing private or non-political interests” (Rawls 1999c, p. 138). There is, in other words, a kind of principled defeasibility about the content of a properly public reason: “Political liberalism does not try to fix public reason once and for all in the form of one favoured political conception of justice. That would not be a sensible approach” (ibid., p. 142).
In other words, the proper content of public reason itself should ultimately be regarded as properly subject to public deliberation and debate, as provisional, as open to question. S

 Says Tully in this vein: “Politics is the type of game in which the framework – the rules of the game – can come up for deliberation and amendment in the course of the game […] A norm of recognition is thus never final, but questionable. It follows that in a free society, existing norms of mutual recognition should be open to public questioning so these reasons can be heard and considered. They should be open to review and potential renegotiation. Reconciliation is thus not a final end-state but an activity that inevitably will be reactivated from time to time” (Tully 2008, p. 146 and p. 309; see also ibid., p. 296 – 298). This seems to me not only fundamentally right, but fully reconcilable with the wide view of Rawlsian public reason: questions of political legitimacy – questions of how, and by whom, is power exercised – ought not be fixed permanently, for there will always be a legitimacy remainder. And according to the view advanced here, public debate over the proper content of public reason – over deliberative rules and procedures, as well as the norms of recognition and participation, and the implications thereof – will have the effect, however seemingly counter-intuitive, of strengthening the bonds of community. Adds Tully: “What shapes and holds individuals and groups together as ‘citizens’ and ‘peoples’ is not this or that agreement but the free agonistic activities of participation themselves […] Reconciliation should be dialogical” (ibid., p. 147 and p. 301; see also ibid., p. 311 – 314).

Of course, there are limits to the open-endedness or defeasibility of what counts as a properly public reason: as I have tried to show throughout this chapter, certain values – freedom, equality, mutual respect and toleration, to name the most prominent – are not subject to public political debate. But it is clearly Rawls’s view that disagreement over what the application of shared justice principles require can be quite deep: debates about what the principles of justice require in terms of public policy and institutional design will lead to significant debate and to the need for reflective political judgment. We may all agree about the importance of self-respect (as a

\[104\] It is important to note that argumentation about the proper content of public reason is importantly different from argumentation about the principles of justice themselves: as stated earlier, the fact remains that the basic structure principles themselves are not up for debate. The kinds of public arguments that Rawls describes in “The Idea of Public Reason Revisited” are arguments about the most faithful, accurate rendering of the guiding values and ideals of those principles into a publicly accessible language.
basic psychological need, a basic primary good), but there are sure to be quite radical disagreements over how this value is effectively satisfied. See, for example, this important passage from *Political Liberalism*: “Accepting the idea of public reason and its principle of legitimacy emphatically does not mean accepting a particular liberal conception of justice *down to the last details* of the principles defining its content. We may differ about these principles and still agree in accepting a conception’s more general features” (PL, §VI.4.5 italics added; see also Laden 2001, p. 16).

In the end, there is no disagreement between Rawls and Waldron when it comes to the question of political debate and disagreement (even though Waldron intends for the following passage to be a critique of Rawls): “Full-blooded disagreement about justice remains the most striking condition of our own politics” (Waldron 1994, p. 385). For Rawls, public debate and deliberation emphatically do *not* have a determinate, indefeasible policy outcome. But what Rawls wants to show us is that the “full-blooded disagreement” of which Waldron speaks, while real and a vital source of energy and vigour, is premised on deeper foundation of agreement on the kinds of principles that ought to the kind the political institutions of the plural community. And inclusive public debate gives us the opportunity to get a better handle on what these principles require in actual political practice; hence, the importance of open public debate in the process of reconciliation.

It is clearly Rawls’s view that religion has an important role to play in this process: Catholic views of the common good and solidarity, for example, can contribute to a refined, and yet more expansive, interpretation of a political conception of justice capable of securing an overlapping consensus (see e.g. Rawls 199c, p. 174). Rawls does recognize the role of religion (and of other comprehensive doctrines) as an important tool in the process of community-building; he does *not* think of religion exclusively as a destabilizing political force, as a phenomenon that undermines the possibility of stability for the right reasons (despite the historical origin of *Political Liberalism* in the Wars of Religion): “In endorsing a constitutional democratic regime, a religious doctrine may say that such are the limits God sets to our liberty; a nonreligious doctrine will express itself otherwise. But in either case, these doctrines formulate in different way how liberty of conscience and the principle of toleration can cohere with equal justice for all citizens.
in a reasonable democratic society” (ibid., p. 151). Rawls then offers as an example the compatibility of Islam with liberal democracy: “the earlier Mecca interpretation of Shari’a supports equality of men and women, and complete freedom of choice in matters of faith and religion, both of which are in accordance with the constitutional principle of equality before the law” (ibid.; see also An-Na’im 1990, pp. 52 – 57).

In a similar spirit, Joshua Cohen offers the example of post-Vatican II Catholicism (the quoted passages within this passage are from an Encyclical by Pope John Paul II on moral theology): “With the conviction that Christ is ‘the way, the truth, and the life’ and, resisting the temptation of ‘detaching human freedom from its essential constitutive relationship to truth,’ I will hold that the truth of the (political) proposition – that individuals have a basic right to religious freedom – follows from the truth of an account of human dignity founded on the doctrine that human beings, created in God’s image, are bound by His laws” (ibid., p. 18; see also Habermas 1998a, §1, Macedo 1998, p. 65 and Weithman 2002, p. 91). Of course, there are prominent strains of Protestantism and Judaism, among others, that endorse liberal principles of justice and modes of public inquiry (Hertzberg forthcoming, Ch. 2).

The important point, in the context of this discussion of religion and public reason, is that “philosophy [too] has good reasons to be open to learning from religious traditions” (Habermas 1998a, p. 109; see also ibid., §4, Habermas 1998b, p. 131, Cooke 2007, p. 225 – 226 and p. 230 and Chambers 2007, §III). Says Habermas in this vein: “The cross-fertilization of Christianity and Greek metaphysics […] promoted an assimilation of genuine Christian elements by philosophy […] Although philosophy transformed the original religious meanings of [responsibility, autonomy and justification, history and remembrance, among others], it did not

---

105 See also Cohen 2009, p. 6: “A liberal political perspective can derive support from other philosophical starting points, including religious views that downplay autonomous choice in favour of substantively correct decisions, regard tradition as a deposit of evolving insight, are metaphysically and epistemologically ambitious, and see religious commitments as providing less latitude and great density of demands.”

106 In his excellent forthcoming dissertation (esp. Ch. 2, Sec. IV.1), Hertzberg dispels the widely held notion among Rawls’s critics that religious obligations will often conflict with the demands of political liberal citizenship: he considers a number of examples of religious views that (a) could plausibly exist in the well-ordered liberal society and that (b) are fully reconcilable with a reasonable political conception of justice – with something resembling Rawls’s first principle of justice (i.e. the most extensive scheme of equal basic liberties) and with “a liberal mode of inquiry that enables free, public and reasoned discussion of political questions” (i.e. public reason, in the Rawlsian sense).
deflate them and exhaust their meaning. The translation of the theological doctrine of creation in God’s image into the idea of the equal and unconditional dignity of all human beings constitutes one such translation” (ibid., p. 110 italics added; see also Habermas 1998b, p. 138 – 139 and p. 142). This is a very important passage for two reasons: (1) it rightly highlights the way in which religion and philosophy grew up together, learning from, and assimilating, each other’s insights and (2) it emphasizes that this process of mutual learning is not complete: philosophy still has much to learn from religion, and so ought not exclude religious claims – even provisionally – from public debate (Chambers 2007, p. 216 – 217; see also Tully 2008, p. 305 and p. 311). In other words, an ethos of mutual respect – based on the possibility of mutual enlightenment – ought to inform the interactions of religious and non-religious citizens. And such an ethos is incompatible with the radically secular worldview that regards religion as nothing more than an archaic holdover from pre-modern societies. Clearly, Rawls concurs with this view of the (potentially) symbiotic relationship between religion and public reason: he often makes reference to Lincoln and the Abolitionists, and to King and the Civil Rights Movement, both of which represent instances of religious beliefs and language illuminating and expanding the political

107 For example, religion has an important role to play in combating “global capitalism’s triumphal march […] It prevents the denizens of the modern secular societies from being overwhelmed by the all-encompassing demands of vocational life and worldly success. It offers a much-needed dimension of otherness: The religious values of love, community, and godliness help to offset the global dominance of competitiveness, acquisitiveness and manipulation that predominate in the vocational sphere” (Wolin 2005, p. 16 – 18). Religion also has a role to play in helping both philosophers and laypersons articulate their (moral) discomfort with human cloning: “Because he is the God of creation and redemption in one, this Creator does not have to work by natural laws like an engineer, or in accordance with the rules of a code like an information scientist. The voice of God that creates life communicates from the outset within a morally sensitive universe. This is why God can ‘determine’ man in the sense that at the same time he makes man capable of freedom and obliges him to be free. Now - we do not have to believe in the theological premises in order to understand the logic that a completely different dependence, presented as being causal, would be involved if the difference assumed in the concept of creation were to disappear and a peer were to take God’s place - if, for example, a person were to intervene with his own preferences into the chance combination of his parents’ sets of chromosomes without not actually having to assume a consensus at least with the other person concerned. Would not the first person who determines another person as he wishes in his natural essence also destroy those same freedoms which exist among equals in order to guarantee their difference” (Habermas 2003, p. 114; see also Cooke 2007, p. 225)? In other words, “there are always likely to be some moral obligations that transcend the language of philosophy” (Chambers 2007, p. 221; see also Cook 2006, p. 195).

108 Such a view is rightly attributed, I think, to Rorty: “As social justice increases, we [secularists] hope there will be less temptation for the poor to murder the rich, and consequently less need for religion as a device for diminishing social unrest, and less temptation to hope for pie in the sky” (Rorty 2003, p. 142; see also Cooke 2007, p. 227). Again, such a view is incompatible with the kind of social co-operation that ought to characterize the Rawlsian milieu.
conception of justice and, by extension, the idea of public reason (see e.g. PL §VI.8.5 and Rawls 1999c, p. 174).

Let us conclude this section by attempting to answer an important question that we raised earlier: does Rawls have an answer to his critics, according to whom a given citizen’s comprehensive commitments – the truth of those commitments – ought to the sole ground of his or her adherence to the laws and policies of the state (Hollenbach 1993, p. 889, Scheffler 1994, p. 13 – 14 and p. 16, McCarthy 1994, p. 51 – 54 and Beiner 2010, p. 286)? Before answering this question, it is important to see the way in which such an objection misses the point of the idea of public reason: the expectation of complete convergence between one’s comprehensive doctrine and the prevailing political conception of justice (which is the subject of an overlapping consensus) is unreasonable in the plural, comprehensively heterogeneous milieu. In such cases, the legitimacy of the liberal state is much too precarious: the consent of various sectarian groups will be withdrawn as soon as complete convergence no longer exists. Indeed, the main thrust of the idea of public reason (as an element of the more demanding morality of robust reasonableness) is to show that one’s religious (and other comprehensive) commitments ought not be the sole basis of one’s adherence to the state: citizens derive meaning from living in, and contributing to, the comprehensively plural (or heterogeneous) community (Rawls 1999c, pp. 171 – 172; see also Habermas 1998b, p. 129). This is perhaps the fullest, most complete expression of the demands of robust reasonableness.

And so, it’s clear that Rawls does not have an answer to those aforementioned (primarily religious) critics, but that his silence on the matter is a principled one: while it may be the case that institutional actors are indifferent to the internal (or psychological) basis of a given citizen’s adherence to the liberal state – what, in other words, is the outward difference between adherence to the state as representative of one’s comprehensive values (which happen to be compatible with the values and language of public reason) and adherence to the state as representative of an inclusive political conception of justice? – political liberalism is precisely the attempt to move beyond a modus vivendi towards stability for the right (principled) reasons (Rawls 1999c, §3). An essential element of the latter condition is the recognition of individual citizens that social life in the plural milieu will require potentially very difficult sacrifices: citizens must recognize the limits of their comprehensive commitments as an appropriate guide for political life. This is yet another indication of the potentially onerous demands of robust
reasonableness: citizenship in the Rawlsian milieu demands that citizens with comprehensive commitments must recognize the insuperable limits of those commitments when it comes to the public, principled justification of political institutions and policies. Of course, the distance between the two will depend upon the nature of those comprehensive commitments; but the important point, given our interest in questions of socialization, is that a constitutive aspiration of political liberalism is to contribute to the emergence of a condition in which private and public values are not only reconcilable but mutually supporting.

In the end, Rawls’s political philosophy of active, committed citizenship – of the willingness to live with a certain amount of (cognitive) dissonance between the private and public realms, and to derive meaning from the management of the psychological complexity that is a by-product of dedicated citizenship in a condition of deep, irreconcilable pluralism – remains in my view a very attractive image of contemporary political and social life. Indeed, we are left with no other choice in the radically plural milieu: what Rawls’s religious critics fail to recognize – this, recall, is the central insight upon which political liberalism is based – is that no two comprehensive doctrines can achieve a condition of complete unity between the existential and political domains. It therefore seems to be the case that Rawls’s critics do not take the fact of deep comprehensive pluralism seriously enough; in the end, the notion of an existence guided entirely by faith fails to take seriously enough the social and political issues that are likely to arise in the comprehensively plural community.

Conclusion.

When it comes to questions about the stability of liberalism under the conditions of plural modernity, Rawls believes that the most important variable is the nature of the comprehensive commitments of citizens. When citizens regard civic virtue as a quintessential expression of those comprehensive commitments – put differently, when citizens’ private beliefs require activity in the service of the common good – liberal institutions are legitimate and stable. Rousseau, by contrast, thinks that the rise of sectarian interest is something to be lamented precisely because of its incompatibility with civic virtue: the existence of plural (and potentially irreconcilable) private moralities compromises citizens’ commitment to the preservation and vitality of the political community. Rawls disagrees: liberalism is stable, on Rawls’s later
account, precisely because liberal institutions have effected a salutary transformation of those
(previously illiberal) comprehensive doctrines with which they have come into contact.
Liberalism is not only compatible with the various (reasonable) comprehensive doctrines present
in society; it is an embedded part – a constitutive dimension – of citizens’ various
conceptualizations of self-interest.

This is what stability for the right reasons means: that, over time, institutional arrangements
transform the character, sentiments and values of those living under them, and that this
transformation has occurred in a liberal direction. Our commitment to institutions is not because
those institutions lead to (contingently) advantageous outcomes for us – this state of affairs
describes a modus vivendi, and is something to be overcome. Instead, where stability for the
rights reasons exists, our commitment to coercive institutions is the by-product of our
commitment to the underlying principles that inform institutional arrangements.

The state of affairs to which Political Liberalism implicitly (but undeniably) subscribes to, then,
is one in which the plural comprehensive commitments of citizens are deeply, thoroughly
compatible with, and provide motivational support for, the guiding values and principles of
justice as fairness. This is the desirable outcome of the history of liberal democracy into which
Rawls inserts his own thought in the Introduction to PL (Rawls 1993, pp. xxv-xxvii). Indeed, for
Rawls, the affective basis of our attachment to the general will is the reasonable comprehensive
doctrine itself; we are committed to the general will, that is, because this commitment is regarded
as an essential expression of our even deeper (affective) commitment to our comprehensive
doctrine; our commitment to our particular good (or way of life) thus enables our adherence to
the general will. What Rousseau and Rawls both see clearly, then, is that individuals must be moved by love in order to recognize, and to be guided by, their rational interests. But Rawls
replaces the vigorous patriotic morality characteristic of Rousseau’s political oeuvre with various
reasonable comprehensive doctrines; again, those reasonable doctrines function in Rawls’s
system as the affective basis of our attachment to the common good of political society.

In the end, the well-ordered community of justice as fairness is best characterized by a powerful,
pervasive ethos of commitment, trust, co-operation, mutual respect and even love (i.e. the
legitimate amour-propre of The Law of Peoples). While liberalism may have been born in the
aftermath of religious war – and it may have had a long period of instability following its birth –
the doctrine itself contributed to the emergence of a political and social condition in which
toleration, co-operation and mutual respect can overcome the enmity previously produced by
comprehensive pluralism. Liberal principles and institutions have effected a major ideological
shift (towards reasonableness) in those comprehensive doctrines that once posed an existential
threat to liberalism itself. This transformation is, to reiterate, the fundamental precondition of
political liberalism (as one candidate conception for our shared theory of justice).

In this chapter, I hope to have shown that Rawls’s image of the well-ordered, stable, solidaristic
political community is a potent resource for rethinking some of the constraints on what is
permissibly said in the public political sphere. Indeed, the Rawlsian public sphere ought to be as
inclusive and discursively permissive as possible. The “wide” view of public reason is a very
potent resource for rebutting those complaints of Rawls’s deliberative democratic critics. I hope
that the preceding arguments have significantly effaced the theoretical distance between Rawls
and those critics. And thus, our re-evaluation of Rawls’s normative theory in light of its under-
appreciated Rousseauian and Hegelian heritage, which is now complete, need not lead to the
wholesale rejection of political liberalism; rather, the amendments proposed are properly thought
of as internal to political liberalism, just as the amendments to justice as fairness are best
understood as internal revisions to that earlier doctrine.
Conclusion

It has been the purpose of the preceding chapters to show that Rawls laments the kind of society in which there are no collective goals, the kind of society in which the state exists only to serve the particular and private ends of its individual members. “Participating in the life of a well-ordered society,” says Rawls in this spirit, “is a great good” (TJ, §86). This is the key aspect of the Rousseauian and Hegelian heritage of justice as fairness, and the centrality of this ideal is preserved in the later articulation of that doctrine, political liberalism. What Rawls’s normative theory aims to avoid, in other words, like Rousseau’s and like Hegel’s, is the view of society as fully private, a society lacking any sense that its institutions possess intrinsic, as opposed to instrumental, significance. Political labour is thought of by the members of such a private society as a burdensome imposition: no one takes account of the good of others, or of what they possess; rather, everyone prefers the most efficient scheme that gives him or her the largest share of assets, resources, power, etc. (TJ, §79). Clearly, Rawls regards this private account of political and social life as a tragedy. Instead, he wants to show that human beings do in fact have shared final ends and that they value their common institutions and activities as good in themselves. The purpose of the state, on this view, is not as an instrumental one: the state’s purpose, instead, is an ethical one – it makes the complete realization of both freedom and self-respect possible, and this is a crucial basis of citizens’ attachment to it.

Commentators have for the most part neglected this Rawls, and what I have tried to show throughout is that this is the by-product of Rawls’s too close affiliation with Kant. The story about Kant is only a part of the story about Rawls. And it is only when we put Rawls in conversation with Rousseau and with Hegel that this new image – an image of Rawls as deeply concerned with the alienating, atomistic tendencies of modern social life – emerges fully. And Rawls’s concern with the good of community is at the centre of his entire normative oeuvre – the Rousseauian and Hegelian dimensions of Rawls’s thought are in no way displaced by the time of Political Liberalism and The Law of Peoples. Indeed, as I have tried to show throughout Chapters 4 and 5, the conspicuous demandingness of robust reasonableness only comes into full
view when we turn to those late texts: the demands of reason become robust when citizens confront the radical pluralism of both their domestic and of the international milieu.

In what sense exactly is robust reasonableness a more demanding ideal of citizenship? And why should we accept, rather than be fearful of, this more demanding ideal? Robust reasonableness, recall, is animated by an imperative of radical inclusiveness: the demands of reasonableness explicitly require the promotion of a maximally inclusive community and, by extension, making each member of the political community feel like a meaningful part of a common project. Simply put, robust reasonableness cannot tolerate exclusion and denied recognition. To feel at home in the political world – to be reconciled to this set of coercive institutions – is to recognize one’s institutional milieu as the source of predictable and genuine recognition and, by extension, self-respect. Of course, the real test of robust reasonableness comes under conditions of radical pluralism: the public sphere must be open to all, for every citizen ought to be in equal and inviolable possession of a share of the public coercive power. This is the liberal principle of legitimacy that informs and animates the labours of engaged citizens: every citizen must see their interests and values reflected in the public political culture. And this ideal extends to the international domain as well: well-ordered peoples are under an obligation of morality to ensure that all peoples everywhere are able to achieve a condition of meaningful self-determination. Robust reasonableness is thus characterized by the psychologically complex interplay of activity and restraint: citizens must actively seek to make the social and political world a home for all, but in certain instances – when citizens confront radical pluralism, both domestic and international – this will also require a restraining attitude of mutual respect. Robust reasonableness demands meaningful recognition, not begrudging or belittling acknowledgment.

In the remaining sections of this conclusion, then, I hope to rescue Rawls from G.A. Cohen’s *Rescuing Justice and Equality*. Cohen is one of Rawls’s most important and most provocative critics: despite his obvious admiration for the institutional structure outlined by Rawls in *Theory*, Cohen is relentless in his view that individual citizens do not subscribe to the quite demanding normative principles characteristic of Rawlsian liberalism (Cohen 1992; Cohen 1997; and Cohen 2008). Cohen thinks of the society of justice as fairness as a fully private one. The society of justice as fairness, in other words, lacks the social cohesion – the sense of fraternity – necessary to bring to life the rigorous morality of Rawls’s normative theory. I disagree. The morality of robust reasonableness *is* the social ethos that Cohen cannot find in Rawls’s normative theory;
robust reasonableness, which is a constitutive dimension of the sense of justice, fully captures individual citizens’ widespread desire to contribute to the well-ordered political community of justice as fairness.\textsuperscript{109} What Rawls describes throughout his entire normative oeuvre are citizens who care deeply about justice; and the demands of justice are conspicuously strenuous. As we shall see below, such a view of the demands of liberal democratic citizenship has important implications on the approach to distributive justice. And so, Cohen has been an implicit interlocutor of the preceding dissertation: my own excavation of the Rousseauian and Hegelian heritage of justice as fairness (and of the morality of robust reasonableness) is, in the end, meant to rebut Cohen’s view that Rawls’s political theory fails precisely because it pays insufficient attention to the necessary other-regarding dimensions of political life. Against Cohen, I have attempted to show that the morality of robust reasonableness is a more attractive because more demanding account of what citizenship in the Rawlsian milieu requires. Let us turn to the distributive implications of such a view.

I.

As the title indicates, in Rescuing Justice and Equality, there are two fundamental points of disagreement between Cohen’s view and (what he calls) “Rawlsian liberal thought.” First, on Cohen’s view, distributive justice ought not tolerate deep economic inequality – citizens’ material prospects should, as a requirement of justice, be roughly equal. This is Cohen’s rescue of equality from Rawlsian distributive justice, which on Cohen’s view leads to excessively inegalitarian outcomes. Second, again on Cohen’s view, the meta-ethical concept of justice

\textsuperscript{109} I take this definition of “ethos” from Michael G. Titelbaum’s “What Would a Rawlsian Ethos of Justice Look Like?”: “I think we should understand an individual ethos as a motive that drives the members of the just society to act on certain principles” (Titelbaum 2008, 303). But the account of Rawls presented below is importantly different from Titelbaum’s account of a full Rawlsian ethos: according to Titelbaum, “the ethos is that which goes beyond the sense of justice […] one can describe the result of such an addition an ‘extended sense of justice’” (ibid, 303f). On my own view, the sense of justice itself – which, of course, comprises robust reasonableness as one of its constitutive elements – is productive of the full social ethos required for the maintenance of just institutions. My own view also differs from Titelbaum’s in that the latter account makes very little mention of the socializing or pedagogical function of institutions; as I shall argue below – and, of course, this is merely a reiteration of earlier claims – the sense of justice itself is the outcome of living under justice institutions. Despite its many merits, this historical dimension, which is such an integral part of my Rousseauian and Hegelian reading of Rawls’s political philosophy, is entirely missing from Titelbaum’s nevertheless important article, which emphasizes the emergence of the necessary social ethos in the original position (see e.g. ibid., 306).
ought not make concessions to practical constraints, which limit the extent to which justice can be applied – justice needs to be distinguished from the practical rules of social living. This is Cohen’s rescue of justice from Rawlsian constructivism (i.e. the original position as a device for the fair selection of social rules). Cohen’s own philosophical methodology investigates “the shape, and the logical implications of, of our deepest normative convictions” rather than the “best-all-things-considered rules of regulation” (*RJE*, 7 – 8). What is the connection between the rescue of equality and the rescue of justice? To the extent that justice is distinguished from other values (i.e. pragmatic concerns for the constraints posed by social life, such as stability, a healthy respect for Pareto optimality and certain forms of publicity), the view that justice requires equality is strengthened, because values other than justice tend against equality. A concern for the practical constraints of social life, in other words, tends to compromise equality (understood in a material sense).

Let us focus on the question of distributive justice first: What, exactly, is the nature of Cohen’s disagreement with Rawlsian distributive justice? Ultimately, it boils down to an issue of incentives: Rawls’s difference principle – as emblematic of the general Rawlsian approach towards social justice – acknowledges that some talented (lucky) people are capable of producing (i.e. earning) more than others, and that it is right for them to be richer than others if the less fortunate are made better off as a result. In other words, the difference principle provides the necessary incentives for those talented people to work and to earn (unequally) despite eventual redistribution. The problem, for Cohen, though, is that these incentives, and ensuing attitudes and actions, run counter to the spirit of the difference principle itself: those talented people “would not need special incentives [to accumulate] if they were themselves unambivalently committed to [the difference principle]” (*RJE*, 32; see also ibid. 74 and 171 and Cohen 1997, 3-4, 9 and 15). There is, in other words, a kind of psychological disconnect between pursing the opportunity to earn unequally and the provision of the necessary conditions for meaningful

110 It is worth noting, here, that this criticism is related to a wider complaint voiced in *RJE*: that Rawls is not engaged in ideal theory at all (i.e. that the need for incentives falls under the domain of non-ideal theory). We take up this critique in detail in Part II below.
equality (i.e. autonomy).\textsuperscript{111} Again: the incentives aspect of the difference principle dissociates the act of accumulation from the moral (i.e. autonomy-guaranteeing) purpose it is meant to serve.\textsuperscript{112}

This is connected to what Cohen calls “The Two Standpoints” in his Introduction to Rescuing Justice and Equality. On Rawls’s view, the difference principle establishes a kind of “moral division of labour”: the (impersonal) state sees to justice (i.e. redistribution according to stated criteria), while citizens pursue the imperatives of his or her personal life within the justice framework of state institutions. According to Cohen, however, this division of labour does not stand up to scrutiny: the (supposedly impersonal) demands of distributive justice (which, on Rawls’s view, are supposed to be restricted to the institutional domain) do in fact (and should) reach into the domain of personal decision-making. Individuals cannot (and should not) ignore the difference principle in their everyday lives and in their everyday economic decisions: the difference principle (not self-interested resource-maximization) should control citizens’ expectations about remuneration. In other words, society requires an “ethos” of justice, not merely rules of justice.\textsuperscript{113} It follows for Cohen that no genuine sense of community can be maintained in the face of the psychological dualism created by this lax interpretation of the difference principle and the economic inequalities permitted by it. Again, Cohen regards the incentives-oriented difference principle as irreconcilable with the sense of justice (as it is described by Rawls himself), even though such incentives may be expedient (not just) in the capitalist cultural milieu. Cohen-ian justice is thus a comprehensive ideal: it is partly institutional, partly dispositional, and neither dimension of justice can be rightly excluded (Cohen 1997, 12 and 26).

\begin{itemize}
\item \textsuperscript{111} Recall that the Rawlsian sense of justice is precisely the identification with, and commitment to, shared justice principles. Cohen is therefore critical of Rawls precisely because the incentives dimension of the difference principle cannot be reconciled with the sense of justice thus understood.
\item \textsuperscript{112} This act of dissociation has one further negative consequence: it makes accumulation (in the name of eventual redistribution) contingent on the desires and intentions of the talented. After all, if talented agents do not find themselves attached to the idea of justice implicit in the difference principle, it is easy to imagine them withholding their wealth-creating (i.e. poverty-ameliorating) labour in the event, say, of an increase in the rate of income tax (Cohen 1997, 16).
\item \textsuperscript{113} Cohen also conceptualizes this critique as the “Basic Structure Objection”: justice cannot be obtained, contra Rawls, by exclusively structural means (Cohen 1997, 9).
\end{itemize}
Cohen’s purpose, then, is to restore the initial liberal emphasis on equality, and to think through the distributive implications of preserving the centrality of the equality principle (cf. his philosophical methodology). Says Cohen: “A maximizing ethos is not a necessary feature of a society, and that, to the extent that such an ethos prevails, satisfaction of the difference principle is prejudiced” (RJE, 143). In the end, justice is fully served on Cohen’s view only if people’s access to desirable conditions of life is equal (within the constraint of a reasonable personal prerogative). Of course, this raises an obvious question: how can we reconcile equality with freedom (of occupational choice)? After all, if everyone ends up with the same economic bundle, there is no impetus to take on burdensome labour (in terms of time or degree of difficulty); equality of outcomes might require that we force the talented into socially useful or productive occupations – into medicine rather than gardening, say – without any concomitant economic rewards (or perhaps the doctor makes more, but her income is subsequently taxed).

---

114 See also Chambers 2006, 83 – 84 for an account of the quite radical demands of the difference principle. Ultimately, Rawls endorses what he calls “property-owning democracy,” which he distinguishes from “welfare-state capitalism” (see e.g. Rawls 2001, Part IV). Says Chambers of this important distinction: “Welfare capitalism redresses the inequalities produced by the basic structure; property-owning democracy offers a redesigned basic structure to ensure minimum or only justifiable inequalities in outcomes. Rawls is interested in a system that minimizes the need for redistribution” (Chambers 2006, 83; see also Krouse and McPherson 1988, 99 and 103). According to Chambers, though, the difference principle ultimate recedes in importance in Rawls’s thought precisely because it is in disequilibrium with the prevailing public political culture. Instead, Rawls’s focus shifts to the question of stability, i.e. the attempt to demonstrate the reconcilability of his first principle of justice with various reasonable comprehensive doctrines. Adds Chambers: “Perhaps if we lived in a world governed by something like the difference principle we would have to think about the sorts of choices that would be required to maintain it. But we are not there yet” (Chambers 2006, 87).

115 See also Titelbaum 2008, 304 – 305 for an account of the ways in which the lexically prior first principle of justice places important limits on the difference principle. I think that Titelbaum is exactly right here: Rawls would only endorse a social ethos compatible with both principles of justice; in other words, the radical material egalitarianism (and concomitant ethos) of G.A. Cohen is likely irreconcilable with a Rawlsian sense of justice, which is also characterized by a higher-order interest in rational autonomy, i.e. the free conditions of self-development. Says Titelbaum of this important point: “Strict reciprocal advantage would sometimes be compromised by the priority of the basic liberties […] the government is prevented by the first principle of justice from assigning individuals to particular occupations for the sake of economic efficiency” (ibid., 312 – 313 and 317). See also PL, VIII.9 esp. p. 335. This is not to undermine the radicalness of Rawls’s difference principle, which is still rightly thought of as a quite demanding principle of redistribution (especially when compared to prevailing practice); instead, it shows that Rawls is not quite as radical as Cohen, and that the Rawls’s relative conservatism saves him from potential violations of the liberties of (talented) individuals. Of course, the Cohen-ian rejoinder is obvious: that those talented individuals who fully identify with the radical (material) egalitarian ethos outlined in Rescuing Justice and Equality would not experience such constraints as a violation of their liberty – quite the contrary: maximiminizing activity (as determined by some central agency, perhaps) is a constitutive aspect of self-realization. I think, in the end, the Rawlsian position makes certain concessions to “men as they are” – ineluctably self-interested, at least to a certain extent – while simultaneously trying to extend the reflective equilibrium of societies characterized primarily by self-interested resource-maximizing behavior. I explore this point in detail.
For Cohen, the essential precondition for this egalitarian outcome is the proper political (ethical) culture (not exogenous regulation): people need to genuinely believe in equality and concomitantly reject morally arbitrary advantages. The reward of medicine, for example, is its social utility rather than personal economic gain even though the doctor will not live an economically or socially deprived life; this is a constitutive egalitarian premise. There is, for Cohen, an importantly social dimension to self-realization.

If the basis of Cohen’s critique is his belief that there ought to be meaningful overlap between the normative orientation of (distributive) institutions and the values and motivations of individual decision-makers, then I submit that there is little disagreement between him and Rawls. In the end, Cohen’s critique gains its traction only by emphasizing the insufficient socializing capacity of the basic structure: his critique of Rawls depends, that is, on a permanent, insuperable schism between institutional arrangements and individual morality. But the starting premise of the revised image of Rawls presented in earlier chapters is that there is no way to distinguish the basic structure from the effects it has on the values, preferences and choices – the morality – of citizens: the site of distributive justice – and of justice more generally – is everywhere, even though the difference principle itself is a principle for institutions. Says Rawls in this vein: “If the so-called private sphere is alleged to be a space exempt from justice, then there is no such thing” (Rawls 1999c, 161).

In other words, Cohen’s account of the insuperable distance between institutional arrangements and the prevailing social ethos misses a foundational principle of Rawls’s normative oeuvre.

---

116 The argument advanced in the Conclusion is importantly different from the argument advanced by David Estlund in his “Liberalism, Equality and Fraternity in Cohen’s Critique of Rawls.” According to Estlund, the disagreement between Rawls and Cohen is not very deep, since both thinkers acknowledge the necessity of some reasonable matrix of incentives in any scheme of social cooperation (Estlund 1998, 112). Instead, my argument is much closer to the one advanced by Joshua Cohen in his “Taking People as They Are?” which emphasizes not the necessity of incentives but the capacity of well-ordered institutions to combat the kind of self-seeking behavior that G.A. Cohen finds so objectionable. See below.

117 See also Titelbaum 2008, 307: “It is psychologically implausible to imagine individuals’ being motivated by one set of principles in their official capacities as citizens and by a conflicting set of principles in their personal lives.”
Indeed, the present dissertation is best characterized by the attempt to show that Cohen’s view is alien to the *Rousseauian* and *Hegelian* spirit of Rawls’s moral and political philosophy; the obduracy of individual conviction to the effects of principle-guided institutions is precisely the Kantian view of morality that the present dissertation has tried to rebut. Again, it seems to me that Cohen ignores Rawls’s important claim that the sense of justice (i.e. the *genuine* commitment to justice principles) is the *outcome* of living under the principles of justice, including the difference principle.\(^{118}\) The following passage from *Political Liberalism* captures this sentiment perfectly and is therefore worth quoting at length: “An economic regime [...] is not only an institutional scheme for satisfying existing desires and aspirations but a way of *fashioning desires and aspirations in the future*. More generally, the basic structure shapes the way the social system produces and reproduces over time a certain form of culture shared by persons with certain conceptions of their good” (*PL*, §VII.7 italics added; see also Rawls 1999b, 7 and J. Cohen 2001, 384).

On Rawls’s view, then, the principles of justice need only apply to the basic structure precisely because of the profound effects it has on the values, choices and motivations of affected citizens. The rules of a just basic structure, in other words, cannot help but transform the morality of individuals subject to those rules. To use the economic terminology employed by Cohen, a maximizing ethos cannot remain impervious to the normative and social pressure of maximinimizing legislation; on the Rawlsian view, the former is almost always thoroughly transformed by the latter (cf. Cohen 1997, 23 – 24; see also J. Cohen 2001, 377 – 378 and 380 – 381). Indeed, as I have emphasized throughout the preceding dissertation, human nature on Rawls’s view is *permissive* and develops in light of institutional circumstances. Rawls does *not* endorse an image of human nature, contra Cohen, that is naturally and inexorably self-interested; hence the ability of institutions to affect dispositional outcomes. Although *today* we might find ourselves in a self-seeking, acquisitive milieu – here, Cohen would emphasize the disconnect between principled institutional structures and individual attitudes, motives and choices – we accept the possibility of the gradual future adoption of the social or cooperative ethos implicit in the difference principle. This claim should be evocative of the Introduction of *Political Liberalism*.

\(^{118}\) Here, as so often, I follow Joshua Cohen: “Rawls shares with Rousseau the thesis that institutions make a large difference to ethos” (Cohen 2001, 376).
Liberalism: although toleration was initially regarded as modus vivendi, it came to be seen as a moral principle worthy of the public, principled assent of citizens. Political institutions, in other words, are always on the vanguard of progressive moral revolutions.

Structures and choices may start out at opposite ends of the motivational spectrum, and then over time begin to coincide precisely because structures influence choices. In other words, institutional justice produces the social ethos that initially supplements it, but which eventually gains a transformative power over it: the kind of society described forty years ago in Theory is still coming into being, and the difference principle is part of Rawls’s attempt to extend – no, expand – the reflective equilibrium of the public political culture characteristic of egalitarian liberal democracies (see e.g. TJ, §17 and §29; see also Chambers 2006, 87). This important point is worth re-iterating: against Cohen – whose view attributes to the society of justice as fairness a strange and unrealistic static quality – my own presentation of Rawls’s thought implies not only that institutional structures influence individual morality and action; the choices and (moral) preferences of individuals can also have a transformative effect on institutional arrangements. Says Rawls in this vein: “[The natural duty of justice] requires us to support and to comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established” (TJ, §19 italics added). In the end, Cohen’s critique attributes to Rawls’s normative thought an artificial (and seemingly insuperable) barrier between institutions and individual morality: such a distinction simply is not present in Rawls’s thought. In fact, the overlooked dimension of Rawls’s moral and political philosophy that this dissertation has tried to draw out is its belief in the essential inextricability of institutions and individual morality, which operate in a kind of permanent dialectic (see e.g. J. Cohen 2001, 378). This is the essential core of the Rousseauian and Hegelian interpretation of justice as fairness that I have advanced in the preceding chapters.

II.

Let us shift our attention from the distributive (equality rescuing) question to the meta-ethical (justice rescuing) question. Here, Cohen pithily summarizes his own view: “I have moved toward the view that justice is justice, whether or not it is possible to achieve it, and that to conform our conception of justice to what is achievable creates distortions in our thought and
also in our practice” (RJE, 155). Rawls is clearly guilty of what Cohen might call an “achievability bias”: the difference principle is best understood as a concession to those talented citizens who expect to be better compensated than their fellows – they think that they need unequalizing incentives in order to produce to the full extent of their talents. This disposition, while unjust (vis-à-vis the difference principle), is a fact about human nature, and is included by Rawls (wrongly, says Cohen) in his thinking about justice. Cohen’s reply: “The flesh may be weak, but one should not make a principle out of that […] I question the idea that a society can be described as ‘truly just’ simply because its unjust people bind themselves against the consequences of their own dispositions to injustice” (RJE, 173-4). Hence, for Cohen, justice is rightly conceptualized as fact-insensitive: if the disposition to injustice happens to be a fact about the world, so much the worse for the world, which is now seen to be incapable of realizing pure justice (see, for example, RJE, 302).

Ultimately, then, the basis of Cohen’s disagreement with Rawls is the excessively fact-sensitive character of the original position: Rawls suffers from a kind of “fact-idolatry.” On this view, Rawls’s thought obscures the distinction between rules of regulation and justice: constructivism (as a device for the selection of social rules and regulations under special conditions of self-interestedness, risk-aversion and limited information) asks ‘what principles should we adopt?’ which is not the same question as, ‘what are our fundamental convictions about justice?’ On Cohen’s view, an answer to the first question presupposes an answer to the second: we ought to conceptualize regulative principles in light of our fundamental (fact-insensitive, pure) normative convictions, not with exclusive reference to contingent facts about human behaviour in this or that social-economic milieu (i.e. with reference to feasibility or practicality) (RJE, 277, 283 and 307). Again, this distinction is obscured – rather, it is purposefully rejected – by Rawls: he conflates fundamental justice principles with social rules. In other words, Rawls doesn’t recognize that justice as fairness is in fact based on trade-offs with other values, such as economic efficiency and human infirmity. Now, trade-offs with other values and attitudes might, in the end, be necessary or practical, but we ought not call that justice.119 Hence, for Cohen, justice as fairness is a kind of disfigured justice, a variant of justice that makes unfortunate

119 Constructivism might be the best way to select social rules, but those rules are not justice.
concessions to – it includes as a part of justice – the kind of self-seeking behaviour that compromises equality. Justice must therefore be rescued from Rawlsian constructivism, which cannot see this distinction.

Let me be clear: with none of this would Rawls agree. Unlike the social ethos question, where there is considerable (although mostly overlooked) agreement, trying to efface the distance between Cohen and Rawls on the meta-ethical question of the nature of justice is not a viable interpretive strategy. For Rawls, the idea of justice is rightly interested in rules for the co-existence of beings like us, in circumstances like ours: the starting premise of the realistic utopia, recall from Chapters 3 and 4, is “men as they are” (see e.g. Rawls 1999b, 45 and TJ, §29). And so, rather than dismiss the circumstances of justice altogether – Cohen, on such a view, detracts from the value of his text by doing so – we have to argue instead about which facts are relevant to considerations of justice; here, comprehensive pluralism and ensuing questions of stability are obvious examples.

Political philosophy, by extension, is rightly attentive to the principles and values (and ensuing attitudes) already implicit in this or that institutional milieu: Rawls is a thoroughgoing Hegelian on this matter. And so, the task of political philosophy contra Cohen – we ought not detach ourselves from the world altogether – is therefore two-fold: (1) to show that the principles that inform our political institutions (the difference principle among them) are rationally designed and that they do, by extension, serve an ethical function, one that is worthy of our dedication and (2) to bring out those instances where attitudes, choices and outcomes are irreconcilable with said principles. Both of these imperatives are captured by the following passage from the Lectures on the History of Political Philosophy: “Political philosophy may try to calm our frustration and rage against our society and its history by showing us the way in which its institutions, when properly understood, from a philosophical point of view, are rational, and developed over time as they did to attain their present, rational form. When political philosophy acts in this role, it must guard against the danger of being a defence of an unjust and unworthy status quo” (Rawls 2007, 7).
III.

Our reconsideration of Rawls’s normative theory in light of its under-appreciated Rousseauian and Hegelian heritage is now complete. And I hope to have presented a compelling case that there is a continuous line of thought between Rawls and certain communitarian strains of the French and German Enlightenments. Rawls is not the thoroughgoing Kantian he is so often portrayed to be. He conscientiously identifies with the Rousseauian and Hegelian goal of overcoming the pervasive atomism of modern market society. Any compelling theory of justice, so this line of thinking goes, must account for the good of community, and must articulate a compelling, and demanding, morality of citizenship, a morality dedicated to institutional preservation and vitality. Citizens, that is, ought to think of the political community as the proper object of reconciliation. This is the heart of the Hegelian heritage of justice as fairness. But, like Hegel, Rawls is committed to the view that only certain regimes are the proper objects of citizens’ reconciliatory labours. This is where we confront the Rousseauian heritage of justice as fairness: only those institutional arrangements that secure the freedom and self-respect of affected citizens are the proper object of reconciliation. This is the intersection of the Rousseauian and Hegelian heritage of justice as fairness, and the confluence of these concerns is also the location of what I have called robust reasonableness. This new, fuller, more attractive account of the demands of the reasonable is the preceding dissertation’s original contribution to our understanding of Rawls’s political philosophy.
Bibliography


