



Swansea University
Prifysgol Abertawe



Swansea University E-Theses

Is justice as fairness a realistic utopia? A critical examination of Rawls's idea of overlapping consensus.

Felton, Maria Paz C

How to cite:

Felton, Maria Paz C (2006) *Is justice as fairness a realistic utopia? A critical examination of Rawls's idea of overlapping consensus..* thesis, Swansea University.
<http://cronfa.swan.ac.uk/Record/cronfa42460>

Use policy:

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

**Is Justice as Fairness a Realistic Utopia?
A Critical Examination of Rawls's Idea of
Overlapping Consensus.**

Maria Paz C. Felton

Thesis submitted for the degree of

Ph.D.

University of Wales, Swansea

Department of Politics and International Relations

2006



ProQuest Number: 10798168

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10798168

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

SUMMARY

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.....(candidate)

Date.....21/03/06.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the corrections is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed.....(candidate)

Date.....21/03/06.....

STATEMENT 2

I hereby consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed.....(candidate)

Date.....21/03/06.....

Summary

Is Justice as Fairness a Realistic Utopia? A Critical Examination of Rawls's Idea of Overlapping Consensus.

The purpose of this thesis is to develop an alternative account of how justice as fairness can be delivered. My general proposition is that if justice as fairness is to be regarded as a realistic utopia, some alterations need to be made into the argument about the possibilities of its realization offered by Rawls. The main problem comes from the idea of overlapping consensus and its association with stability. A crucial working conjecture is that we need overlapping consensus neither to stabilize justice as fairness, nor to justify it. My alterations to Rawls's argument also originate as a result of two theoretical burdens within political liberalism: its notion of stability and its account of the fact of reasonable pluralism. Both ideas have been inadequately problematized, and therefore demand significant revision. The idea of stability needs to be revised because of its exclusive focus on questions of moral coherence at the expense of more relevant aspects such as institutional and political stability. The account of reasonable pluralism also needs to be revised because it is both too narrow and significantly unfocused. A crucial consequence that follows from these revisions is that, in order to preserve institutional stability under the conditions of pluralism, we need to render the stability of institutions separately from the coherence of citizens' overall moralities. These two conceptual revisions lead me to effect three alterations into Rawls's argument. In particular I show that: 1) constitutional consensus is necessary, but not yet sufficient, for stability of the required kind, 2) political justification is necessary and sufficient for the right kind of justification, and 3) overlapping consensus is part of moral, not political, justification and, therefore, not required for a political project such as Rawls's.

ACKNOWLEDGEMENTS

Thanks are due to various people and, first of all, to my supervisor, Dr. Mark Evans. I am particularly grateful for his support and for helping me in the pursuit of something that I have always valued -clarity of thought. I wish to thank him also for his comments on both general and detailed matters. I am also especially indebted to Professor Bruce Haddock for his encouragement and for the numerous discussions that we had about political philosophy, and to Professor Rex Martin for his invaluable help with the reading of Rawls's texts. I am also grateful to the Department of Political Theory and International Relations at the University of Wales, Swansea for the support offered during the course of my study.

My work on this thesis would not have been possible without the help and encouragement of my husband, John. He has taken over, particularly in the last year, some of the tasks that the running of a household with three young children normally requires. I am deeply thankful to him for sharing some of the worry that goes with writing a thesis. My mother has also been involved in this thesis by providing invaluable childcare at crucial times. I am also grateful to her. My three children, Antonio, James and Daniel, all born whilst studying for this Ph.D have been a constant source of motivation.

I finally wish to thank the audiences of three conferences at the Universities of Manchester and Essex for various valuable comments on some of the ideas contained in this thesis, and to the editors of *Politeia* for publishing my paper 'Overlapping Consensus and Critical Justification'.

TABLE OF CONTENTS

INTRODUCTION	1
1 THE BACKGROUND	10
1. Justice as Fairness: Political not Metaphysical	11
2. The Idea of an Overlapping Consensus	16
3. The Domain of the Political and Overlapping Consensus	27
4. Political Liberalism and Beyond	32
5. Conclusion	42
2 OVERLAPPING CONSENSUS: A CRITICAL ENQUIRY	44
1. Critiques of Overlapping Consensus	44
2. Overlapping Consensus: A Critical assessment	66
3. Conclusion	89
3 THE PROBLEM OF STABILITY	91
1. Stability: Some General Considerations	92
2. Stability in TJ	105
3. Stability in PL	135
4. Conclusion	141
4 PLURALISM AND POLITICAL LIBERALISM	144
1. Reasonable Pluralism: Features and Implications for Political Liberalism	147
2. Problems with Rawls's account	153
5 CONSENSUS AND STABILITY	173
1. Constitutional Consensus: Preliminary Remarks	174
2. Rawls on Constitutional Consensus	182

3.Stability and Constitutional Consensus	189
6 CONSENSUS AND JUSTIFICATION	202
1.The limits of Constitutional Consensus	202
2.Rawls on Justification	208
3.The Idea of Public Reason	213
4.Public Reason and Political Justification	218
5.Stability and Legitimacy	223
7 OVERLAPPING CONSENSUS REVISITED	226
1.Overlapping Consensus: Realistic or Unrealistic Utopia?	227
2.Overlapping Consensus: A Dystopia?	235
3.Overlapping Consensus and Critical Justification	239
CONCLUSION	245
BIBLIOGRAPHY	247

Abbreviations

<i>A Theory of Justice</i>	TJ
'Justice as Fairness: Political not Metaphysical'	PnM
'The Idea of an Overlapping Consensus'	IOC
'The Domain of the Political and Overlapping Consensus'	DPOC
<i>Political Liberalism</i>	PL
'The Idea of Public Reason Revisited'	IPRR
'The Law of Peoples'	LoP
<i>Justice as Fairness: A Restatement</i>	JFR

INTRODUCTION

The purpose of my thesis is to develop an alternative account -to the one offered by Rawls himself- of how justice as fairness can be delivered. In doing so, my general intention is not to challenge the overall aim of his project, but to improve on his particular framework. The initial impetus for pursuing this programme was prompted by an increasing dissatisfaction with the Rawlsian notion of overlapping consensus and its association with the problem of stability. Whilst I was generally sympathetic towards his attempt to reconcile people's political perspectives with their overall moralities, the more I looked into the notion of overlapping consensus, the more I was led to think that it was too utopian, and therefore not a realistic utopia. By way of clarification, let me state at the outset what constitutes, in Rawls's own terms, a realistic utopia. A given project of political philosophy can be regarded as a realistic utopia if, by stretching the possibilities of actual feasibility, it fulfils the key role of reconciling ourselves to the political and social conditions of our times.¹ In the case of constitutional democracies (the case that I will be considering in this thesis), the relevant political and social conditions are defined by the fact of reasonable pluralism.

This problem of utopianism was something that Rawls himself had considered, and he offered a solution based on a historical developmental account; a solution that I judged to be far from adequate. I then began to think that an account of how justice as fairness is possible perhaps did not require an overlapping consensus after all, and this is how this thesis took shape. My proposition is thus that if justice as fairness is to be regarded as a realistic utopia, some alterations need to be made into the argument about the possibilities of its realization offered by Rawls. A crucial working

¹ Rawls, J., *The Law of Peoples with 'The Idea of Public Reason Revisited'* (Cambridge, MA: Harvard University Press, 1999), p.11.

conjecture here is that we need overlapping consensus neither to stabilize justice as fairness, nor to justify it.

These alterations also originate as a result of what I generally regard as two theoretical burdens within political liberalism: its notion of stability and its account of the fact of reasonable pluralism. Perhaps I need to make clear at the outset that here I am not disputing the legitimacy of Rawls's concern with stability and his view that some conception of this notion is required to answer the question of how justice as fairness is possible. Neither am I suggesting that some version of pluralism should not drive the direction of his project. What I am claiming is that these two notions -that are so crucial to the project- have been, in my view, inadequately problematized, and therefore both demand a significant revision in their conceptualisation.

The conceptualization of stability needs to be revised principally because of its exclusive focus on the question regarding the coherence between citizens' political and 'internal' outlooks. This coherentist account based on the concordance of fit between peoples' political and comprehensive views prioritises moral stability over and above other kinds of stability. Moral stability is attained when the majority of active citizens of a particular society have internalised the political values that are embedded in their society's basic institutions and have incorporated them into their respective domains of values. However, there remains, as we shall see, other questions of stability still unresolved, for instance, questions about political and institutional stability. A polity can be said to be institutionally stable when its citizens agree to adopt a particular set of institutions as a result of a process in which they have participated and on the basis of reasons that they can all accept. Such a polity may also achieve political stability if its political institutions, in their day-to-day running, do not depart from the values that ground those very institutions. Whereas Rawls's idea is

that we should concentrate on moral stability as both necessary and sufficient for the attainment of the two other kinds of stability (that is, or political and institutional stability), I argue that this position is untenable. And, it is untenable because the fact of pluralism makes moral stability not only implausible, but also undesirable.

My suggestion that the conceptualisation of stability given in *Political Liberalism* (hereafter PL) is a theoretical burden demands some explanation. Concerns of stability arise because of Rawls's preoccupation with part III of *A Theory of Justice* (hereafter TJ), in particular with the tension between the account of stability presented there and the pluralism contained even in societies well ordered by justice as fairness. The notion of overlapping consensus is designed precisely to overcome this tension. However, even though Rawls has offered a different solution to the problem of stability (than the one given in part III of TJ), the fact remains that both accounts of stability are given as intended solutions to a problem issued in the same way. All there has been is a transposition of the problem from TJ to PL without appropriate regard for the important differences between the two projects. The fundamental question is this: if in TJ arguments for justice as fairness, including the arguments for its stability, are conducted at a comprehensive level, it is not unreasonable to think that such arguments may be motivated by concerns that arise from those comprehensive arguments themselves. Hence, to borrow a distinction from Rawls, whereas the concept of stability cannot be but one, its conceptions may vary as a result of the separate sets of questions that each book addresses.²

Rawls's account of reasonable pluralism also needs to be revised primarily because it is both too narrow and significantly unfocused. It is too narrow because it centres, almost exclusively, around doctrinal conflict, particularly on religious, moral

² In straightforward terms the main question of PL is the one that this thesis aims to address, namely, how can justice as fairness be possible. The separate question of which is the best conception of justice is implicit in the project of TJ. This is a question that can only be pursued at a comprehensive level.

and philosophical disagreement. Furthermore, the account is unfocused because it does not pay sufficient attention to conflicts of a political nature. Since pluralism is one of the things that make a political order unstable, if the account of pluralism is both too narrow and significantly unfocused, then overlapping consensus must fail as a solution to stability. A political project such as political liberalism should, above all, be concerned with political pluralism. Hence, a more complex picture of pluralism emerges, one that reflects more adequately the conditions of modern democracies.

The crucial consequence that follows from these revisions is that in order to preserve institutional stability under the conditions of pluralism, we need to render the stability of institutions separately from the coherence of citizens' overall moralities. As a result of this, having presented my own versions of both stability and pluralism, my next step is to examine the idea of constitutional consensus in the light of my suggested problematization of stability. This constitutes the first alteration to the overall argument about the possibility of justice as fairness. The idea of constitutional consensus is a crucial part of Rawls's solution to the problem of utopianism. My aim here is to abstract from his historical developmental story and to assess constitutional consensus on its own and, in particular, to see whether it can neutralize possible tendencies towards institutional instability. Here I argue that, although constitutional consensus is necessary for this kind of stability, it is not yet a sufficient condition. My argument for the necessity of constitutional consensus also constitutes an argument against the sufficiency of a *modus vivendi* in relation to institutional stability.

Once institutional stability is achieved in a constitutional consensus, we need to consider this hard question: is institutional stability put under strain by adding to the constitution in accordance with a conception of justice like justice as fairness? We need to examine whether broadening the constitution might produce the undesired

effect of a return to institutional instability. This seems to have been Rawls's fear and his main motivation behind overlapping consensus. But the answer given here is somewhat different. Rawls's answer is that the problem of stability requires moral justification of a certain kind. This thesis argues that solving stability only necessitates political justification or, to put it in the vocabulary of political liberalism, freestanding justification. This constitutes the second alteration in the general argument for the possibility of justice as fairness. To put it clearly, freestanding justification is necessary for the right kind of stability, and both necessary and sufficient for the right kind of justification.

A possible objection that might be raised against my argument is the following: 'If your concern is merely institutional stability why, given the importance that you attach to the fact of pluralism, don't you just stop at constitutional consensus?' In other words, if pluralism is a good thing, why not, as pluralists generally insist, stay within a constitutional consensus? My reply to this is that a constitutional consensus says very little about justice. A constitutional consensus merely reflects the idea of a democratic minimum. Hence, as a minimum, it leaves important political questions unanswered. A constitutional consensus might have been sufficient for stability if justice were not our concern. But, within a constitutional consensus, injustices remain. I do not need to go into the full catalogue of injustices present in constitutional democracies for the reader to grasp how we, as political philosophers, may be motivated, even by our own particular comprehensive doctrines, to map out how our own conceptions of justice may be realized. This is a project worth pursuing so long as we are prepared to develop political justifications for our own understandings of justice and their requirements. This thesis, however, is only concerned with the particular case of justice as fairness.

The third and final alteration to Rawls's argument is, as it were, an alteration by default. Thus, I will examine the possibilities of overlapping consensus in the knowledge that it is unable to stabilize justice as fairness. Nonetheless, I venture into its scrutiny in order to see whether there is anything there that we can preserve. My general conclusion is that overlapping consensus is part of moral, not political justification and, as such, not required for a political project such as Rawls's.

Hence, to reiterate, this thesis provides:

1. An examination, critique and assessment of the central idea of overlapping consensus.
2. Two conceptual revisions -those of the notions of stability and reasonable pluralism.
3. Three alterations to the argument offered by Rawls of how justice as fairness is possible.

In the development of this project, my thesis falls clearly into three parts. The first two chapters deal with the idea of overlapping consensus. The first chapter looks at this idea from both a historical and a textual perspective. By historical I mean that my incursion into overlapping consensus focuses on how the idea was developed over approximately an eleven-year period. Hence my examination starts with the 1985 paper 'Justice as Fairness: Political not Metaphysical' and concludes with the 1996 edition of PL. By examining these writings closely, I aim to provide a balanced exposition of the notion under study. After that I turn, in chapter 2, to an analysis of overlapping consensus. In the first part of the chapter, I survey five different objections that critics have raised against this idea, whilst, in the second part, I assess those objections. My aim here is to be as favourable to overlapping consensus as possible. However, this kind of 'defensive' assessment can only partially succeed as

there remain two critical problems. These problems concern the link between consensus and stability, and between consensus and justification. The rest of my thesis is, in a sense, an attempt to address such crucial difficulties.

The following two chapters are devoted to my two conceptual revisions. Chapter 3 is concerned with the problem of stability and this examination takes us back to TJ. I try to trace the origins of this very Rawlsian preoccupation with the purpose of establishing whether his initial concerns can be identically transposed to his new framework, particularly given the non-comprehensive character of his latter work. I argue that, given the different aims of the two books, the problematization of stability of TJ cannot just be transferred to PL, and that the latter requires a different conceptualisation of stability, one that takes primordial account of political and institutional aspects. The second of my conceptual revisions is effected in chapter 4. This chapter attempts to sketch out Rawls's idea of reasonable pluralism and to point to its shortcomings. Two related ideas are also discussed here: the idea of reasonableness and the notion of the burdens of judgement. Such close examination of these three ideas leads me to offer a version of pluralism centred around the political and one that makes use of a much weaker conception of reasonableness.

The third part of my thesis contains the three specific alterations to the argument for the possibility of justice as fairness. Chapter 5 considers the idea of constitutional consensus in the light of my revised account of stability. In examining this idea I will be showing how a high degree of stability is achieved there. A constitutional consensus –a consensus on certain institutional procedures that goes beyond a *modus vivendi*- is, I argue, a necessary, but not yet sufficient, condition for institutional stability. Institutional stability also requires, as I will show in chapter 6, a political, or freestanding, justification of justice as fairness. This is a necessary step in

the argument for two reasons: 1) because constitutional consensus on its own can scarcely address matters of justice, and 2) because of the likely risks that the process of broadening the constitution may create for institutional stability.

The thesis concludes by revisiting the troubled notion of overlapping consensus. In particular, I survey four senses in which overlapping consensus might be concerned with issues of justification. I discuss whether, and to what extent, overlapping consensus can be considered a realistic utopia, an unrealistic utopia, a dystopia, or whether its concern lies with a process of critical moral justification. I conclude by showing that, although the notion of overlapping consensus might have a place in moral theory, it is too utopian for the project of political liberalism.

To conclude these introductory remarks, I wish to mention two important reasons why I consider this project of improving on the Rawlsian framework to be worthwhile. One of the reasons derives from Rawls's own views of the purposes of political philosophy. According to Rawls, an important task for those who engage with the questions and problems of political philosophy is that of 'probing the limits of practical political possibility.'³ As I also share this conception of the purposes and goals of political philosophy, I decided to test Rawls's project of political liberalism and, in particular, the idea of overlapping consensus against his own aims and aspirations. This exercise caused me to make the alterations to his argument mentioned above. In a sense, I am concerned with the internal overall coherence of Rawls's latter work.

There is another important reason why I have engaged in this project. This can be explained as follows. There is nothing inconsistent in claiming that whereas political philosophers may possess more or less comprehensive conceptions of what

³ Rawls, J., *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001), p.4.

justice entails, they should nonetheless provide political justifications for those conceptions. I do not think that this separation has been adequately stressed either by Rawls or by his critics. But, in my view, this is crucial to our appraisal of the main differences between TJ and PL. Hence, I see the aim of offering a freestanding political justification for a more or less self-contained and, arguably, comprehensive conception of justice, as the central aspect that drives the programme of PL. And, it is this aspect that is key to our understanding of the possibilities of justice as fairness. Thus, my project does not arise as a result of a defeat. It is not motivated by a realization that, if overlapping consensus is too demanding, we may just forget about it and account for the feasibility of justice as fairness in some other way. This would make my thesis vulnerable to the charge of following a 'sour grapes' methodology. By contrast, my aim is to show how there are crucial reasons, and reasons that are internal to political liberalism itself, that make overlapping consensus unsuitable for a project of political justification. Any project designed to flesh out the conditions that a conception of justice -comprehensive or otherwise- must fulfil for it to be politically justified, must be worth pursuing.

CHAPTER 1

THE BACKGROUND

We start our journey by examining the key idea of overlapping consensus. This idea has been the object of most of the critiques raised against the Rawlsian move to the 'political'. These critiques will be considered in some detail in the following chapter but, before I do so, I wish to offer a plausible account of how the concept itself was developed. Hence this first chapter concerns itself with some exegesis of Rawls's texts. This is not, I hope to show, an otiose exercise of Rawlsian stratigraphy but, rather, an attempt to identify the different sets of questions that prompted Rawls to coin, define and redefine this idea over a period of roughly twelve years, starting with the publication of 'Justice as Fairness: Political not Metaphysical'¹. This chapter follows a chronological order with almost each section dealing with one key text. The exception is the fourth and final section which focuses on both PL and Rawls's 'Reply to Habermas'.² This first chapter prepares the ground for the second one which is concerned with an examination and evaluation of some significant weaknesses of overlapping consensus.

The idea of overlapping consensus figures prominently in most of Rawls's post-1985 writings.³ It is not my aim here to provide an exhaustive exegesis of each of

¹ 'Justice as Fairness: Political not Metaphysical' (1985), in Rawls, J., *Collected Papers*, edited by Freeman, S. (Cambridge, Mass.: Harvard University Press, 1999), pp.497-528. Hereafter PnM.

² *Political Liberalism* (New York: Columbia University Press, 1993); the revised paperback edition (1996) includes an additional introduction (pp.xxxvii-lxii), and Rawls's 1995 paper 'Reply to Habermas' as Lecture IX (pp.372-434). This was originally published in 'The Journal of Philosophy'. All references are to the 1996 paperback edition.

³ Apart from PnM and PL, the relevant sources for this chapter are 'The Idea of an Overlapping Consensus' (1987), in *ibid.*, pp.421-48, hereafter IOC and 'The Domain of the Political and

the writings that consider overlapping consensus, but to relate it to other key ideas within Rawls's theory and, more importantly, to the main questions and concerns of the Rawlsian project. My tentative suggestion here is that these concerns may have changed, or at least may have been slightly altered, in the route from 'Political not Metaphysical' to PL and beyond. The one thing that is clear is that these writings offer a substantive departure from TJ. Although the idea of overlapping consensus was coined in his first book, it was used in a different context and had a different meaning than the one it has come to have in his post-1985 writings.⁴ The main departure from TJ stems from the fact that Rawls did not distinguish in his first book between a comprehensive moral doctrine and a political conception of justice. This is not the place to speculate whether the conception of justice given in TJ was actually a comprehensive moral doctrine,⁵ but what seems clear is that Rawls thinks that this is a possible and plausible reading of TJ, and of the last part of the book in particular. In what follows I offer a partial (to suit our specific purposes) reconstruction of the development of the notion of overlapping consensus.

1. Justice as Fairness: Political not Metaphysical.

In this first conceptualisation of overlapping consensus Rawls ties his 'new' understanding of justice as fairness to the role of political philosophy in a democratic

Overlapping Consensus' (1989), in *ibid.*, pp.473-96, hereafter DPOC. Other key sources for this thesis are the 1997 paper 'The Idea of Public Reason Revisited' which was later included in *ibid.*, pp.573-615 (all my references are to this later source), hereafter IPPR, *Justice as Fairness: A Restatement*, edited by Kelly, E., (Cambridge, MA: Harvard University Press, 2001), hereafter JFR, and *The Law of Peoples with 'The Idea of Public Reason Revisited'* (Cambridge, MA: Harvard University Press, 1999), hereafter LoP.

⁴ Overlapping consensus is used in *A Theory of Justice* (Oxford: Oxford University Press, 1973), first published in 1971. There is a second, revised edition, published by Harvard University Press in 1996. All references are to the 1973 Oxford University Press text. In TJ the idea of overlapping consensus is closely connected with issues of civil disobedience, see pp.387f.

⁵ Rawls suggests this in his first Introduction to PL. Also, Samuel Freeman has suggested that chapter 9 of TJ pushes justice as fairness towards comprehensiveness in 'Political Liberalism and the Possibility of a Just Democratic Constitution', *Chicago-Kent Law Review*, 69 (1994), pp.619-68, esp. p.628.

society. An appropriate theory of justice for a constitutional democracy should not just avoid resting on controversial philosophical and metaphysical views, but also be independent of such views. Political philosophising under democratic conditions requires the application of the principle of toleration 'to philosophy itself'.⁶ The two main aims of this paper are 1) to show that justice as fairness can in fact be understood as a political, as opposed to a metaphysical conception, without altering much of its content, and 2) to argue for the desirability of political conceptions of justice (as opposed to metaphysical ones) since 'as a practical political matter no general moral conception can provide a publicly recognized basis for a conception of justice in a democratic state'.⁷

That justice as fairness is intended as a political conception can be spelled out in three ways. Firstly, the conception of justice is specifically designed to apply to the basic structure of society. This is captured by the idea of society's basic structure as the subject of justice. In this first sense, the conception of justice presented in TJ was already political. Secondly, justice as fairness is political because the content of justice is taken from the fundamental ideas present in the political institutions of a constitutional democracy. This sense of the political was also present in TJ. Justice as fairness draws on the ideas of liberty, equal citizenship and on the notion of rights to be secured by justice that are fundamental to a constitutional democracy. What TJ did not contemplate, however, was the wider notion of a political tradition. Justice as fairness is intended as a political conception because its fundamental ideals and how they are understood –the public traditions of their interpretation- are taken from the political culture of constitutional democracy. It is in this context that the idea of overlapping consensus appears for the first time. Because justice as fairness arises

⁶ PnM, p.388.

⁷ Ibid., p.390.

‘from within a certain political tradition’⁸, the hope is that it can be supported by an overlapping consensus formed by all the opposing philosophical and religious doctrines that coexist in a democratic society.

The third sense in which justice as fairness can be understood as a political conception is closely related to the role of political philosophy in a democracy. An important task of political philosophy is to identify possible sources of agreement in the midst of political disagreement. The kind of fundamental disagreement that Rawls seems to have in mind is that between competing understandings of the values of freedom and equality (between the ‘liberties of the moderns’ and the ‘liberties of the ancients’)⁹. Hence the task of justice as fairness as a political conception is twofold: firstly, to elaborate principles of justice for a society’s basic institutions in a way that arbitrates between the two traditions, and secondly, to explain the ways in which these principles provide a more adequate understanding of a basic democratic requirement, i.e., the conception of citizens as free and equal. Rawls is confident that his two principles can provide this necessary basis of agreement, and that justice as fairness will pass the test of ‘reflective equilibrium’.¹⁰

Hence we start from a specific political tradition, and try to abstract from disputed moral and religious controversies. The idea behind this is that since political disagreement is sufficiently deep, why burden ourselves with other sources of disagreement? After all, it is not the appropriate role of the state to settle matters of morals or religion: ‘We do this not because [moral or religious] questions are unimportant or regarded with indifference, but because we think them too important and recognize that there is no way to resolve them politically’.¹¹ The conclusion that

⁸ Ibid.

⁹ Ibid., pp.391-92.

¹⁰ Ibid., p.393.

¹¹ Ibid., p.394.

Rawls draws from this is that an adequate basis for agreement under democratic conditions must stay within the political.

Besides overlapping consensus, two other important Rawlsian ideas are modified here to fit in with his move to the political. The idea of the person and the notion of the original position also get a makeover. The conception of the person must also stay within the political; a person is a citizen, a 'fully cooperating member of society over a complete life'¹² with a capacity for a sense of justice and for a conception of the good. This is a political idea designed to serve political purposes, but it is also a normative idea. The idea of the original position also gets recast; it is now to be viewed as a device of representation, as a means of public reflection and clarification. There are no metaphysical or ontological considerations involved in being placed behind the veil of ignorance. The original position is simply a clear way of making vivid our views about social cooperation amongst free and equal citizens.¹³

That justice as fairness is to be understood as a political conception is central to an understanding of liberalism as a political doctrine.¹⁴ The conception of the person as political is particularly important to this understanding. People are not required to view themselves as free and equal all the way down. This might seem strange as the moral ideal of autonomy has been central in the development of liberalism, but 'philosophical toleration' allows people to pursue their incompatible, and often incommensurable, conceptions of the good, provided they accept for themselves, and for others, the principle of free and equal citizenship. Comprehensive liberalisms, like those of Kant and Mill, are not a suitable political basis, but rather, two, amongst many, conceptions of the good. Because justice as fairness is elaborated from the ideas present in the political culture of a democratic society, Rawls is

¹² Ibid., p.397.

¹³ Ibid., pp.399-403.

¹⁴ Ibid., p.408.

confident that it may be affirmed by those who live in this kind of society, regardless of whatever other comprehensive views they may hold. Justice as fairness thus aims 'to identify *the kernel of an overlapping consensus*'¹⁵, the shared set of ideas required to elaborate a political conception of justice. For Rawls, this seems to be both a necessary and sufficient condition of political theorizing under democratic conditions.

The idea of overlapping consensus appears somewhat underdeveloped in this paper, for this reference to the kernel of an overlapping consensus certainly lacks clarity. An unfamiliar reader might be led to think that justice as fairness needs to focus on only those political ideas that are *de facto* shared. Rawls partly clarifies this by introducing another 'new' idea, that of a *modus vivendi*.¹⁶ However, here a *modus vivendi* is not a state of affairs, but a way of endorsing the political conception. Rawls concedes that in an overlapping consensus, some people may affirm the political conception as a *modus vivendi*, guided by either prudential considerations or reasons of self-interest. He nonetheless thinks that he can avoid the 'Hobbesian charge' because overlapping consensus focuses on a moral conception (justice as fairness), and also, because each of the comprehensive doctrines taking part accepts justice as fairness from a moral point of view, each doctrine from within its own reasons.

So far considerations of stability have been absent from the discussion. Rawls devotes only the last section of this paper to explain how liberalism as a political doctrine understands the basis of social unity and stability.¹⁷ The idea of overlapping consensus to some extent defines the basis of social unity, but the question of its stability is determined by the content of the comprehensive doctrines forming the consensus, and of the particular links between justice as fairness and each comprehensive doctrine. There are different ways of balancing the values of justice as

¹⁵ Ibid., p.410.

¹⁶ Ibid., p.411.

¹⁷ Ibid., pp.411-14.

fairness and the values of each doctrine, but this, Rawls admits, is a 'highly speculative' exercise since social conditions to a certain extent influence the range and nature of comprehensive doctrines, and social conditions are in turn influenced by the particular conception of justice adopted by the society in question.¹⁸ The question of stability is not pursued in this paper although in its final footnote, Rawls makes an interesting comment about the incompleteness of part III of TJ.¹⁹ With regards to stability, TJ only considered one case within an overlapping consensus, the case where the public conception of justice is affirmed in itself, without invoking further values, beyond political values. Hence, in TJ the discussion of stability is limited to one case; the footnote suggests that a general account of stability must extend to the remainder of the cases within an overlapping consensus. These other cases are not discussed in PnM.

2.The Idea of an Overlapping Consensus.

We have seen how in Rawls's initial understanding of overlapping consensus, considerations of stability only appear at the margins, and they are mainly connected to the question of the basis of social unity in a democratic society. In this 1987 paper Rawls offers something in the way of a proper definition of stability. I leave aside for now discussion of the features of overlapping consensus since most of the account remains much the same in subsequent writings. What is of interest here are Rawls's comments on the role of political philosophy in a democratic society and the conclusion that he draws from this.²⁰

We can illustrate the nature of his comments in the following way. Different types of societies define the role of political philosophy in their own distinct ways.

¹⁸ Ibid., p.414.

¹⁹ Ibid., note 33.

²⁰ IOC, pp.421-23.

Each society formulates its own political problems, and confers a sense of urgency to a few of them. Political philosophy in a given society focuses on these few questions that people in that particular society perceive as the more relevant and pressing ones. But societies not only have a political history; they also have a history of political theorizing. Hence in modern constitutional democracies political philosophy focuses on questions that relate to the theory and practice of constitutional democracies. Political philosophy, at least in this kind of polity, does not start from scratch, so to speak. We can assume that people, in varying degrees, have some notion of the values embedded in the democratic ideal. Some people may have a somewhat loose grasp of democratic values whilst others may have a more articulated understanding of the basis of a democratic polity. Therefore, given that the role of political philosophy depends 'on the society it addresses', in a constitutional democracy one of its main roles is to work out a *political* conception of justice that is able to provide a shared *public* basis of justification of its political and social institutions, and that is able to guarantee its *long-term stability*. Underwriting this role of political philosophy is the coexistence in a democratic society of a variety of opposing philosophical, religious and moral doctrines, or 'the fact of pluralism'²¹.

The fact of pluralism pushes us in the direction of what Rawls here calls a *regulative* political conception of justice. The political conception helps formulate political principles that reflect the public understanding of the values embedded in the democratic ideal and that are to frame the constitution. A political conception of justice is thus able to perform the first task of political philosophy, the task of uncovering the shared public basis of justification of a democratic society's main institutions and practices. The fact of pluralism and the limits of democratic politics

²¹ Ibid., p.421.

make it necessary that we look for a political conception²². Hence Rawls develops on this aspect that was central to the argument of PnM. Let us look at this in some detail.

For Rawls an appropriate conception of justice for a democratic society needs to be political 'all the way down'. By this he does not mean that it is not a moral conception, but that *it is worked up to apply only* to the basic structure of a society. This means that in the process of elaboration of this conception, the fact that it is going to apply only to the basic structure, and not, for instance, to the general conduct of individuals, needs to be taken into account. Hence Rawls is careful to distinguish between a political conception that is only applicable to the basic structure, and general views like perfectionism, utilitarianism, idealism and Marxism that, although can be applied to the basic structure, may generate other demands on individuals and societies.²³ The motivation behind this is that the fact of pluralism, and more generally the social and historical conditions of a democratic society, go against the adoption of a general and comprehensive view as a public basis for a political conception of justice. The fact of pluralism is not another historical contingency, but the result of the free exercise of reason under favourable circumstances.

Hence Rawls not only isolates the 'political' as the domain in which the conception of justice applies, but also restricts the foundations for justice to the political values embedded in the democratic tradition of thought and practice. This conception, Rawls suggests, will be appropriate to apply at least to constitutional essentials. As a result, agreement on a political conception of justice thus conceived

²² He writes: 'The thesis of the first part of my discussion is that the historical and social conditions of modern democratic society require us to regard a conception of justice for its political institutions in a certain way. Or rather, they require us to do so, if such a conception is to be both practicable and consistent with the limits of democratic politics'. Ibid., p.423.

²³ Ibid., p.424.

will secure the stability of a constitutional democracy. This is his working hypothesis in IOC.²⁴

This is a working hypothesis simply because Rawls himself concedes that an overlapping consensus may not always be possible. An overlapping consensus may require the modification of certain beliefs and values held by actual citizens in democratic societies. But the point is to show that convergence on a political conception of justice is at least a realistic possibility, and that this will secure stability. There is no other alternative given the fact of pluralism since to privilege one comprehensive doctrine at the expense of all others will be self-defeating. If the goal is consensus, and there is widespread disagreement on moral, philosophical and religious long-standing controversies, these controversies need to be somehow bypassed.

Now trying to bypass long-seated disputes does not mean that the political conception is able to avoid comprehensive doctrines altogether. Rawls accepts that this may be necessary to rebut the charge of scepticism or indifference with respect to religious, philosophical or moral truths. It may well be appropriate to invoke general views (provided their long-standing disputed features are avoided) along these lines: 'what is the least that may be asserted; and if it must be asserted, what is its least controversial form?'²⁵ But Rawls seems to go further than this. He admits that it may well be possible and desirable that citizens develop more than one political conception from the shared political ideas, and that these may compete over time for citizens' support. Hence the fact of pluralism is also applicable to conceptions of justice. But it

²⁴ This is clearly summarized in the following passage: 'Since we are concerned with securing the stability of a constitutional regime, and wish to achieve free and willing agreement on a political conception of justice that establishes at least the constitutional essentials, we must find another basis of agreement than that of a general and comprehensive doctrine. And so, as this alternative basis, we look for a political conception of justice that might be supported by an overlapping consensus.' Ibid., p.425.

²⁵ Ibid., p.429.

also means that in the elaboration of those political conceptions, citizens are guided by their own general views provided that these are kept within the political.

Thus in IOC, as we have seen, Rawls introduces the fact of pluralism as an important assumption about the conditions of political philosophy in democratic societies, and also refers for the first time, although without appropriate elaboration, to the notion of constitutional essentials. A third important idea introduced in this paper, albeit rather briefly, is free public reason.²⁶ A conception of justice is not complete unless it specifies the guidelines for public inquiry and assessment of evidence in the application of its two principles. We do not only need to excuse Rawls's brevity here, but also his lack of clarity when he refers to the limits of public reason posed by the fact of pluralism: 'there is (...) no better practicable alternative than to limit ourselves to the shared methods of, and the public knowledge available to, common sense, and the procedures and conclusions of science when these are not controversial.'²⁷

Rawls anticipates some possible objections to the idea of overlapping consensus and hence devotes the second half of this paper to answer four kinds of objections. The first objection expands on an aspect already mentioned in PnM, namely, that an overlapping consensus is not a *modus vivendi*. A *modus vivendi* as defined by Rawls is a consensus on self- or group- interests that arises as a result of political bargaining. The weakness of a *modus vivendi* lies in its dependence on the contingency of different factors like the relative strength of each party to the consensus, the position of their leadership and so on. For these reasons, it is unable to secure the long-term stability of such an agreement.²⁸ To this charge of contingency, Rawls opposes the moral character of a consensus that is moral in its object (the

²⁶ Ibid.

²⁷ Ibid.

²⁸ Ibid., pp.430-31.

political conception of justice), and moral in its ground (it is affirmed on the moral bases provided by comprehensive doctrines). Furthermore, the long-term stability of such consensus is secured by the fact that those taking part in it (those affirming the political conception) will continue to do so as they will remain unaffected by the contingencies in the distribution of political power. Citizens will continue to hold their comprehensive views, and thus to affirm the political conception, even though their views may become dominant. Rawls remains positive that such contingencies will not affect the overlapping consensus since comprehensive doctrines compete for people's moral allegiance, and not for power. Three examples are provided of general views that may belong to an overlapping consensus, although again, as in PnM, the links between these general views and the political conception are not developed. The three views considered are a religious doctrine with an account of free faith, the liberal moral doctrines of Kant and Mill, and a general political doctrine that affirms the political conception on the basis of its own political values. Interestingly, Rawls is skeptical about whether utilitarianism can actually be included given its index of citizens' desires, preferences and interests as these may well go beyond the limits posed by the basic structure. The hope for utilitarianism is that it undergoes a process of self-revision that enables its incorporation into the overlapping consensus.²⁹

The second possible objection advanced by Rawls is that 'the method of avoidance' (the avoidance of general doctrines in thinking about a political conception of justice) expresses indifference or skepticism with regards to the truth of the political conception. For Rawls it is important that skepticism and indifference are rebutted as this will make his project self-defeating at the outset. Unnecessary conflicts with comprehensive doctrines must be minimized. Hence, the method of

²⁹ Ibid., pp.430-34.

avoidance requires that the truth or falsity of general doctrines is neither affirmed nor denied, and that citizens ascribe the same predicate that they ascribe to their comprehensive doctrines (true, reasonable...) to the political conception.³⁰ This predicate-transfer is what gives the political conception its moral ground. However, a further charge is that the retreat to the political and the method of avoidance must express scepticism or indifference to be able to sidestep fundamental philosophical, moral and religious disagreement. Rawls's answer to this is that questions are not removed from the political agenda because they are intractable, but that in looking for a political conception, an important aim is that it allows us to discriminate between the questions that can be removed from the political agenda and those that cannot, even though these may be controversial. For example, the principle of equal liberty of conscience rules out from the political agenda matters concerning religious truth, but it allows questions regarding the separation of church and state.³¹ Rawls nevertheless concedes that a political use of certain aspects of comprehensive doctrines may be necessary to respond to possible fundamental threats to the basics of constitutional democracies. If someone insists on using political power to settle matters concerning salvation, the solution may be to use the doctrine of free faith and its support for the principle of equal liberty of conscience to show how it is possible to hold both a comprehensive religious doctrine and a political conception of justice. In doing this, the aim of consensus needs to be kept firmly in mind so that 'we do not state more of our comprehensive view than we think would advance the quest for consensus.'³² This is required by the limits of free public reason. In fact, even though an agreement that is moral in its object and moral in its ground may be thought of as conducive to truth, this is an unnecessary step and it may reduce the possibilities of obtaining

³⁰ Ibid., p.434.

³¹ Ibid., pp.435-36.

³² Ibid., p.436.

public agreement. This is a task to be conducted by citizens themselves, guided by their own comprehensive doctrines.

The third possible objection to overlapping consensus is that a political conception must in itself be general and comprehensive in order to deal appropriately with the conflicts of justice that are likely to arise in a democratic societies.³³ This view rests on a fundamental mistake. Such mistake is to equate the abstract character of a conception of justice with its comprehensiveness. A conception of justice must be abstract so that it can transcend political conflict and locate itself in the theoretical domain. This, however, does not imply its comprehensiveness. To show how an adequate conception does not need to be comprehensive to be effective, Rawls appeals to his third model case within an overlapping consensus, the case of a political conception of justice that is affirmed on its own grounds. Such conception reflects political values that are in itself sufficient to support a just regime. Not only that, those political values normally outweigh other values that may conflict with them. The citizens affirming this political doctrine may also hold other values, but these are not articulated into a fully comprehensive doctrine. Their reasoning is that provided the basic liberties are guaranteed within the framework of a just constitution, no conflict of values is likely to occur. This example also serves to explain how a political conception expresses values that 'normally outweigh whatever other values (may) conflict with them.'³⁴ Rawls concedes that the most adequate conception of justice for a democratic society is a liberal conception of a general kind. It is liberal because it tries to uncover the basis of social cooperation. This is not possible unless exclusive political values are upheld. In the final instance, the wider the consensus, the weaker will be the conflict. This is why the values of the political conception need

³³ Ibid., pp.437-40.

³⁴ Ibid., p.439.

not be weighed against the values of comprehensive doctrines. Finally, the hope is that through ‘the work of reconciliation by free public reason’ it may be possible to obtain ‘a sufficiently concordant fit among political and other values as displayed in an overlapping consensus.’³⁵

We turn now to the final objection, namely, that an overlapping consensus may be utopian. This will occur if it is either impossible to obtain a consensus, or if the one that is obtained turns out to be unstable.³⁶ The answer that Rawls provides against this charge depends on the plausibility of an interpretative story concerning the development of liberal democracies. It depends on the initial acceptability by a non-democratic society of a liberal conception as a result of a *modus vivendi*, as the only alternative available to political and civil conflict. For Rawls it is certainly plausible to think of a process of allegiance (towards a liberal conception of justice) developing out of that initial acquiescence. Here Rawls relativizes the extent to which endorsement of a political conception depends on the affirmation of a comprehensive doctrine. In theory, a political conception may be either derived from a comprehensive doctrine, or be compatible with it. However, in practice, most people do not consider the links in this way, as this involves highly speculative questions. We must instead think of people as holding partially and loosely articulated comprehensive doctrines, and as affirming at the same time the political conception without actually realizing the possible connections between the two. Such connections may be worked out over time, and the process may involve adjustments and revisions of their own comprehensive doctrines. Rawls’s answer to the question of how a liberal

³⁵ Ibid., p.440.

³⁶ Ibid., pp.440-44.

conception of justice might gain an allegiance for itself turns into an account of how a regulative liberal conception leads to a stable constitutional regime.³⁷

A liberal conception of justice is able to stabilize a constitutional regime because it fulfils three conditions. Firstly, the liberal conception is able to fix, 'once and for all',³⁸ basic rights and liberties and to provide an order of priority amongst them. Here the stabilizing effect is that basic rights and liberties are placed beyond the calculus of partisan interests and beyond circumstantial changes. It restricts unnecessary bargaining and establishes a platform for equal mutual recognition. Secondly, the liberal conception also provides a form of reasoning for settling the political disputes that legitimately remain on the agenda. This is specified by the idea of free public reason. The stabilizing effect of public reason partly relates to its simplicity. Guidelines must be set up in accordance with forms of reasoning generally available to citizens. By contrast, reasoning within comprehensive doctrines usually involves a high degree of theoretical competence not always available to citizens. To offer public reasons is to offer reasons that others are in a position to understand. And thirdly, when the basic rights and liberties are institutionally fixed and citizens are used to giving one another public reasons, the desire to cooperate with others in political life is developed. The three aspects –the fixed basic liberties and their priority, the cooperative political virtues and free public reason- support one another. The basic liberties define the terms of social cooperation, this in turn encourages the political virtues which are strengthened by the use of free public reason. This framework provides a robust stabilizing constraint.³⁹

Hence in answering the possible charge of utopianism, Rawls discusses two interrelated aspects. The first aspect is how a political conception initially endorsed as

³⁷ Ibid., pp.442-44.

³⁸ Ibid., p.442.

³⁹ Ibid., p.444.

a *modus vivendi* might turn out to be the focus of an overlapping consensus. The second aspect relates to the stabilizing mechanisms afforded by a political conception once this conception has institutionally fixed the basic rights and liberties and their priority.

The last two sections of this paper anticipate two aspects of Rawls's theory that are developed in later papers. The first aspect relates to his understanding of the historical and social conditions of modern democracies and the implications of the fact of pluralism.⁴⁰ The fact of pluralism is a permanent fact in democratic societies and can only be suppressed by the oppressive use of power. Rawls considers this to be not only a fact about democracies, but also a restriction on political theorizing. This connects with the second aspect. The conditions of modern democracies make it necessary that liberal philosophy focuses on political liberalism defined as 'the view that under the reasonable favourable conditions that make constitutional democracy possible', claims that 'political institutions satisfying the principles of a liberal conception of justice realize political values and ideals that normally outweigh whatever other values oppose them.'⁴¹ Political liberalism tries to occupy a middle ground between Hobbes, on the one hand, and Kant and Mill on the other. Both are, to a certain extent, unstable. Whereas a Hobbesian agreement lacks moral force, the liberalisms of Kant and Mill are too comprehensive to command widespread support. The hope is that, by taking people as they are, political liberalism is able to maintain a just constitutional regime.⁴²

⁴⁰ Ibid., pp.444-46.

⁴¹ Ibid., p.447.

⁴² See, for instance: 'In exhibiting the possibility of an overlapping consensus in a society with a democratic tradition confronted by the fact of pluralism, political philosophy assumes the role Kant gave to philosophy generally: the defence of reasonable faith. In our case this becomes the defence of reasonable faith in the real possibility of a just constitutional regime.' Ibid., p.448.

3.The Domain of the Political and Overlapping Consensus.

We turn now to the last of the three papers on overlapping consensus prior to the publication of PL. This paper elaborates further on this idea and also introduces the few ‘missing pieces’ of the theory: the two stages in the presentation of justice as fairness and the idea of the burdens of reason.⁴³ Furthermore, Rawls establishes a closer connection between overlapping consensus and stability, and makes some links with TJ.

Partly as a response to critical comments by G.A Cohen and Habermas⁴⁴, Rawls is keen to show that his political conception is not political in the wrong way. The charge here is that concerns of justice appear secondary over the primary concern of reaching a stable consensus. To this Rawls responds by identifying the two stages in the presentation of justice as fairness, the free-standing stage where the political conception is presented in its own terms, and the second stage of overlapping consensus which provides an account of the stability of justice as fairness. He then goes on to explain the background reasons for his turn to the political. These are given by four ‘facts’: the fact of pluralism, the fact that a continuous endorsement of one comprehensive doctrine can only be obtained by the oppressive use of power, the fact that a stable democracy must be supported by a substantial majority of its citizens, and the fact that the political culture of a democracy already contains the intuitive ideas from which to elaborate a suitable political conception of justice.⁴⁵ These are facts of political sociology and human psychology that are, for Rawls, incontestable. The first two are particularly important since the enjoyment of basic rights and liberties under free institutions by itself leads to pluralism. Hence a political

⁴³ For distinction between the two stages, see DPOC, p.474, and for the burdens of reason, later called the burdens of judgement, see *ibid.*, pp.475-8.

⁴⁴ Such critical comments were made in private discussions as Rawls explains in *ibid.*, p.473*n*.

⁴⁵ *Ibid.*, pp.474-5.

community (a society united on a single comprehensive doctrine) can only be maintained by the use of unjustified force, by oppression and, almost certainly, repression. But the fact of pluralism is not, what we may call, a bare sociological fact. Disagreements are not signalled by simple differences in matters on personal interests or tastes. It is not even the case that people can be easily misled or misinformed. The idea of the burdens of reason aims to explain how reasonable disagreement might occur. The disagreements that interest Rawls are disagreements amongst reasonable people, amongst people with the two moral powers who accept the terms of social cooperation. The burdens of reason refer to the possible sources of disagreement amongst reasonable persons, to 'the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgement in the ordinary course of political life.'⁴⁶ The sources of disagreement are not limited to values, but include the sources of evidence, the particularity of judgements, the indeterminacy of concepts and the partiality of experiences. The burdens of judgement underwrite a fifth general fact, namely, that fully reasonable and conscientious people are not likely to agree when making judgements. For Rawls the important feature of these facts is that they are not mere historical contingencies, but they derive from the exercise of reason under free institutions.

The next step is to refine a bit further the political character of his conception of justice. Firstly, the political conception must be linked to an account of its stability at the outset.⁴⁷ A political conception is stable if it is able to generate its own support. This in turn connects with the three features of a political conception, particularly with the second and third: 1) the political conception is a moral conception that is worked up for a political subject, namely, the basic structure of society, 2) endorsing a

⁴⁶ Ibid., p.476.

⁴⁷ Ibid., pp.479-81.

political conception does not entail endorsing a single comprehensive doctrine, and 3)the conception is worked up from the shared fund of ideas that belong to a democracy's political culture. A political conception is not general because it focuses on political values alone, and it does not include other ideas and values that belong to the non-political sphere, as this would supersede the limits of a political liberalism.

Although political liberalism is neither a fully nor a partially comprehensive doctrine, its content is related to the liberal tradition, to basic rights and liberties and their priority. Now the conditions of modern democracies (summarized in the five general facts) that lead to the need to look for a political conception takes Rawls to his key idea of constitutional essentials.⁴⁸ Such essentials are constituted by the basic rights and liberties, the three democratic powers and questions concerning majority rule, political and civil rights and liberties, and the protection of the rule of law. A suitable political conception of justice for a democratic society must, at a minimum, settle these urgent essentials. It is precisely because of this urgency that Rawls thinks that the political conception may be endorsed by comprehensive doctrines. If this task is accomplished, that is in itself sufficient for, as a political conception of justice, it may be of little relevance to specify the nitty-gritty of democratic politics such as the economic and social questions that are normally left to the legislative power.⁴⁹

Rawls then goes on to define the 'political' as a special domain. The political domain is characterized by two main features: 1)the non-voluntary character of political authority and 2)the coercive nature of political power which, in a democracy, is limited by the conditions of citizenship.⁵⁰ The political thus differs from other domains, for instance, the associational (which is voluntary in character) and the personal and familial (based on affection). Thus, not only must a suitable conception

⁴⁸ Ibid., pp.482-4.

⁴⁹ Ibid., p.481.

⁵⁰ Ibid., p.482.

of justice be political, the domain to which the conception is to apply must also be appropriately limited so that 'its main institutions can gain the support of an overlapping consensus.'⁵¹ Hence, justice as fairness is not applied moral philosophy. Its content is restricted to political values, and its domain of application is also limited to the basic structure of a democratic society, and, more precisely, to the essentials of its constitution. Political liberalism does not deny the existence of the wider realm of values, but simply their unsuitability for the domain of the political. Hence, if someone claimed that because the doctrine of *Extra ecclesia nulla salus* is the true one, the state should enforce it, political liberalism's answer is to say that, whereas the doctrine might be true, it is unreasonable to use state power to enforce a view that many citizens do not accept.⁵² Political liberalism tries to avoid possible replies from other comprehensive doctrines, as the best strategy is to stay within the political domain. Political liberalism only claims that enforcement by the state of a comprehensive doctrine is unreasonable, only as a last resort political liberalism may publicly refer to the incorrectness of a doctrine.

Rawls goes on to explain how political liberalism conceives the links between the political conception and the diversity of comprehensive doctrines.⁵³ This explanation is divided in two parts. In the first part, citizens themselves, guided by their liberty of conscience, decide how the political values are connected to their own particular views. The hope is that these values alone provide a suitable basis of political justification. In the second part the focus switches to comprehensive doctrines. Each comprehensive doctrine, as shown by the history of religion and philosophy, is to think of whether its values are congruent with, supportive of, or merely compatible with, the political conception. The diversity of comprehensive

⁵¹ *Ibid.*, p.483.

⁵² *Ibid.*

⁵³ *Ibid.*, pp.484-5.

doctrines means it is necessary to look for an overlapping consensus whereas the reasonableness of such doctrines defines the possibility of its achievement.

Rawls's attention now turns to the second stage in the presentation of justice as fairness. Once a conception of justice is worked up from the shared ideas implicit in the democratic tradition of values, the second stage is to consider how stable such a conception is likely to be. Only a political conception can be thought to be practicable given the fact of pluralism. But we are not to think of stability as a merely practical matter, although this may be important at the outset. What is important is the reasons and forces grounding that stability: the sense of justice it defines so that citizens willingly accept the political conception and its principles, and the reasonableness of citizens who endorse it.⁵⁴ As a liberal conception, justice as fairness is designed to gain the support of citizens who affirm reasonable, but incompatible, comprehensive doctrines. Since justice as fairness is addressed to each citizen's reason, it provides an account of political legitimacy, that is, an account of how political power must be justified to citizens under the conditions of pluralism: 'If justice as fairness were not expressly designed to gain the reasoned support of citizens who affirm reasonable although conflicting comprehensive doctrines (...) it would not be liberal.'⁵⁵ The preceding discussion leads Rawls to explain the ways in which his new ideas may depart from the view given in TJ.⁵⁶ The distinction between a political conception and comprehensive doctrines was not made in TJ. Rawls therefore grants that this might have led to the understanding that justice as fairness as presented there was grounded on a comprehensive doctrine under the name of 'rightness as fairness'.⁵⁷ This reading is further reinforced, Rawls concedes, by his account of the well-ordered society in

⁵⁴ Ibid., pp.486-7.

⁵⁵ Ibid., p.487.

⁵⁶ Ibid., pp.488-90.

⁵⁷ Ibid., p.488.

part III. The account assumed that the citizens in a well-ordered society were required to affirm the same conception of justice and the same comprehensive doctrine from which it derived. This is no longer considered a realistic possibility, and this is what led Rawls in the first place to formulate his idea of overlapping consensus. TJ failed to acknowledge the fact and extent of pluralism so that its account of the well-ordered society now seems utopian. In Rawls's view, his move to the political makes his theory less utopian.

4. Political Liberalism and Beyond.

Despite Rawls's acknowledgement of some possible misgivings with respect to the account of stability given in TJ, it is not until we get to PL that we find Rawls's precise diagnosis of the problem that led him to revise his theory, and therefore an explanation of the differences between his two books.⁵⁸ The main departure comes from the attempt to resolve what he identifies as an 'internal' problem in justice as fairness. This internal problem is defined by an inconsistency between his conception of justice as fairness taken as a whole and the account of its stability. As a consequence of this inconsistency, the idea of the well-ordered society as presented in TJ turned out to be unrealistic and both the account of stability and the idea of a well-ordered society given in TJ were in need of revision. Although in IOC and DPOC he hesitates over the extent to which justice as fairness was presented in TJ as a comprehensive doctrine, he seems now quite prepared to concede that justice as fairness quite clearly appears as at least a partially comprehensive doctrine, and that this is incompatible with the project of political liberalism. The main goals of this project also make the idea of the well-ordered society given in TJ unrealistic. The

⁵⁸ PL, pp.xvii-xix.

diversity of comprehensive doctrines coexisting in a modern democracy makes the endorsement of a conception of justice that is grounded on a particular comprehensive doctrine fairly unlikely. What has forced the changes is, Rawls admits, the problem of stability, or in other words, the question of ‘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?’⁵⁹ This led Rawls to think about the nature, structure and content of a conception (a political conception) of justice able to sustain stable support over time.

As I mentioned above, the features of overlapping consensus remain much the same in PL, and also the two stages in the presentation of justice as fairness. However, Rawls makes a few refinements that give us the complete picture of his project. Firstly, he introduces the liberal principle of legitimacy. This principle relates to the two features of the domain of the political, the non-voluntary character of political authority, and the coercive nature of power. Given these two ‘facts’ of the political and given the nature of democratic citizenship, this principle states that ‘our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason.’⁶⁰ Therefore a democracy’s constitutional essentials must be settled according to this liberal principle of legitimacy, but also, Rawls now adds, the matters of basic justice that may arise in the legislature. Decisions on both kinds of questions must be made exclusively on the basis of political values and the principles and ideals that they express. Political values are sufficiently robust as they frame the very conditions (the favourable conditions) under which we live our social and political

⁵⁹ Ibid., p.xx.

⁶⁰ Ibid., p.137.

lives. The political values defined by justice as fairness are, in the first instance, freestanding. It is then up to citizens themselves to work out how the political values relate to the rest of values that they hold. Historical experience shows that there can be at least compatibility between political and non-political values suitably understood.

The liberal principle of legitimacy connects with Rawls's appraisal of the question of stability. In a constitutional democracy any stabilizing mechanism that a conception of justice may have must not be based on the use of state power in an arbitrary way. This would violate the liberal principle of legitimacy. The answer to the problem of stability has two parts.⁶¹ The first part responds to the question of whether people who have had experiential knowledge of living under just institutions do in fact acquire the necessary sense of justice to comply with those institutions. This is the same question that Rawls asked in ch. 8 of TJ, and his answer remains the same: our basic psychological assumptions about living under just institutions show that citizens are likely to acquire a sense of justice robust enough to provide stability. The second part of the stability question is limited to the case of the just institutions of a democracy with the permanent feature of reasonable pluralism. Stability, under the conditions of reasonable pluralism and its effects (the coexistence of diverse, and incompatible, comprehensive doctrines) requires an overlapping consensus. This is at the very heart of Rawls's political liberalism: '...justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen's reason, as explained within its own framework. Only so is it an account of the legitimacy of political authority as opposed to an account of how those who hold political power can satisfy themselves, and not citizens generally, that they are acting

⁶¹ Ibid., pp.140-4.

properly.⁶² The idea of overlapping consensus is very closely connected to both stability and legitimacy.

The next step is to refine the definition and features of overlapping consensus from the one already given mainly in IOC, but also in DPOC. The first crucial point that Rawls makes is that the doctrines that are a part to the consensus must be *reasonable*. The ground of an overlapping consensus must be defined only by reasonable doctrines with the exclusion of unreasonable and irrational ones. This is because Rawls is now concerned, not with the fact of pluralism as such, but with the fact of reasonable pluralism. Pluralism may well be the outcome of living under free institutions, however, not all exercises of reason are equally valid.⁶³

Now Rawls connects the three features of an overlapping consensus (its moral object, moral ground and stability) to an account of the ways in which it differs from a *modus vivendi* (the first of four possible objections to his idea of overlapping consensus). His views here are much the same as those given in IOC. However, he now refers to the depth and breath of a consensus and to the specificity of its focus, that is, how deep the consensus feeds into citizen's comprehensive doctrines, to which range of institutions it applies and how deep and wide a practicable political conception of justice needs to be. Rawls's aim is to find 'the deepest and widest feasible conception of political justice.'⁶⁴ A more shallow and narrower consensus may well be feasible, but this may not be sufficient as a reasonable basis of social unity. The options are roughly speaking two. One is to find a political conception with the aim of seeking a consensus on just the constitutional essentials, and the other is to find a political conception with the aim of seeking a consensus that covers the basic

⁶² Ibid., pp.143-4.

⁶³ Ibid., pp.58-66.

⁶⁴ Ibid., p.149.

structure as a whole. For Rawls only the second option provides an adequate basis for social unity suitably understood.⁶⁵

Rawls's discussion of the second and third possible difficulties of overlapping consensus –that it is neither indifferent nor sceptical, and that it does not presuppose a general and comprehensive view- does not vary from IOC. He has modified, however, his answer to the fourth objection –the utopian nature of overlapping consensus. Although the problem is cast on the same terms (an overlapping consensus may either be unfeasible or unstable) and his answer also depends on the plausibility of his views on the history of liberal democracies, he now introduces the crucial step of constitutional consensus.⁶⁶ Hence this important idea comes associated with the question of the actual prospects for overlapping consensus. It is introduced as a way of dealing with the question of utopianism. Its role is to serve as a necessary step, but only as a step, in the transition from a *modus vivendi* to an overlapping consensus.

In the first stage Rawls discusses how liberal principles of justice that are initially accepted on the basis of a *modus vivendi* are finally established as constitutional principles. In a constitutional consensus principles are accepted in themselves and are not rooted in particular conceptions of justice and in ideas of the person and society.⁶⁷ A constitutional consensus has shallow grounds and lacks a specific focus. It establishes the required minimum framework (basic principles and democratic electoral procedures) for arbitrating amongst different groups and interests. As it only covers an unspecified class of basic rights and liberties and certain democratic procedures, and does not cover the whole of the basic structure, the

⁶⁵ Ibid., pp.149-50, and 158-168.

⁶⁶ This follows from remarks made by Baier in his assessment of overlapping consensus as a solution to stability. Baier claims that a constitutional consensus, which in his view already exists in advanced constitutional democracies, is more appropriate to solve stability; 'Justice and the Aims of Political Philosophy', *Ethics*, 99 (1989), pp.771-90.

⁶⁷ Ibid., p.164f.

consensus is also narrow. It is interesting to note that Rawls's explanation of how a constitutional consensus may be obtained is almost identical to the explanation offered in IOC of how a *modus vivendi* may turn into an overlapping consensus.⁶⁸ It seems therefore that Rawls might have thought that there was too big a gap between the stabilizing constraints of a *modus vivendi* and those provided by an overlapping consensus. So the explanation given in IOC of how a *modus vivendi* derives into an overlapping consensus now becomes the explanation of how a *modus vivendi* turns into a constitutional consensus.

The steps to an overlapping consensus are defined by the processes of deepening, widening and focusing of the constitutional consensus. In the process of deepening, citizens and political groups are led to articulate their principles into more or less elaborated political conceptions so that they can explain to others in the public forum their preferred policies. These conceptions provide a more complete basis from which political debates may be conducted since 'a constitutional consensus at the level of principles viewed apart from any underlying conception of society and citizen (...) is a consensus taken literally. It lacks the conceptual resources to guide how the constitution should be amended and interpreted.'⁶⁹ Political conceptions are also necessary for an effective exercise of the powers of the judiciary, including the powers to accept or reject legislative enactments. With regards to breadth, a consensus including only a few essentials is insufficient to deal with the variety of important issues -issues that form the business of democratic politics- that border on constitutional essentials and matters of basic justice. Hence, in the process of widening the consensus, citizens and political groups are expected to come up with political conceptions that cover the whole of the basic structure, and that contain the

⁶⁸ IOC, pp.440-2.

⁶⁹ PL, p.165.

resources to deal with such matters as civil and political rights and liberties, and welfare rights. Finally, in the process of focusing, the important question relates to the range of views that can be worked up from the ideas present in a democracy's political culture. Justice as fairness is just one interpretation of those ideas, but there may be others. In any case such a political conception 'would certainly be typical of the focal class of an overlapping consensus.'⁷⁰ Rawls admits that different views about social and economic interests are bound to result in the elaboration of different liberal views. Hence, the narrower the difference between those views, the narrower will be the range of political conceptions in an overlapping consensus. But if the different liberal views elaborated from the democratic political tradition, instead of narrowing the agreement, actually encourage political and economic conflict, it may not be possible after all to obtain an overlapping consensus.

Rawls concludes Lecture IV of PL by discussing how the political conception and the diversity of comprehensive doctrines may be related.⁷¹ His purpose is to show that the particular realm of values of the various reasonable comprehensive doctrines either support, are congruent with, or at least do not conflict with the political values that underlie the political conception of justice. To the three cases discussed in IOC and DPOC (comprehensive liberalism, the religious view with an account of free faith and the pluralist view) he now adds a fourth example, the utilitarianism of Bentham and Sidwick that he discusses instead of Mill's liberalism. The kinds of relationships between the political conception and comprehensive doctrines that he has in mind are of the following form:

1.The political conception and its principles can be derived via deduction from the Kantian moral ideal of autonomy.

⁷⁰ Ibid., p.167.

⁷¹ Ibid., pp.168-72.

2. Utilitarians who follow Bentham and Sidwick (but possibly utilitarians of other persuasions as well) could be seen as accepting the political conception as the one that best approximates to the principle of utility given the conditions of knowledge in modern societies marked by the fact of uncertainty.

3. The comprehensive pluralist view already contains the political values within its wider realm of values. The political conception can be endorsed by a balancing of judgements that prioritises the political over non-political values.

4. A religious doctrine with an account of free faith will accept the political conception on the basis of the principle of toleration since it underpins the fundamental basis (the basic liberties) of a constitutional regime.

Rather surprisingly, Rawls does not really work out these links in any rigorous detail. Moreover, he does not think that this is necessary as it may lead to a fair amount of argumentative speculation. This aspect has been the subject of much criticism, so I will return to it in chapter 2. Ultimately, Rawls claims, an overlapping consensus 'achieves compliance by a concordant fit between the political conception and the comprehensive views together with the public recognition of the great values of the political virtues.'⁷²

We have almost concluded our exegesis but there are two remaining aspects that we need to mention. The first one is only a minor point. In the Introduction of the 1996 paperback edition to PL, we can find a more precise diagnosis of what was in need of revision in TJ. In this second introduction to PL Rawls finally admits that the conception of justice presented in TJ was biased towards the comprehensive liberalism of Kant.⁷³ The conception of the well-ordered society of TJ was such that citizens' sense of justice was defined by a Kantian comprehensive doctrine. Hence it

⁷² Ibid., p.171.

⁷³ Ibid., p.xlii.

was the idea of the well-ordered society presented in part III of TJ that led him to make some substantial revisions in order to account for the fact of reasonable pluralism. This in turn required, as we have seen, a reformulation of his conception of justice and of its main ideas as political ideas.

The second remaining aspect is more substantial as it bears on a key question, namely, the role that overlapping consensus plays within Rawls's theory. This is interesting because in Rawls's discussion (which constitutes his reply to a criticism made by Habermas concerning the function of overlapping consensus) considerations about justification are added to those of stability.⁷⁴ Habermas had asked whether the comprehensive doctrines in the consensus had a justificatory role (if they performed the function of strengthening and deepening the justificatory basis of the political conception) or if their function was merely to secure social stability. A second related question is whether the term reasonable implies a truth claim, or simply a tolerant attitude with respect to comprehensive doctrines. Rawls's answer is to break down the process of justification of the political conception into three stages: 1) *pro tanto* justification, 2) full justification by individual citizens, and 3) public justification by political society.

Pro tanto justification is concerned with the justification of justice as fairness from political values alone. This roughly corresponds to the freestanding stage. The political values that it defines and its public reason are sufficient to provide a reasonable answer at least to questions regarding constitutional essentials and matters of basic justice. However, as this is only a political justification, it may be insufficient as a basis of justification all the way down, when the rest of the realm of values are introduced. The second and third stages of justification both consider how the

⁷⁴ The 'Reply to Habermas', initially published in 1995 in 'The Journal of Philosophy', constitutes Lecture IX in the 1996 edition of PL. The relevant section of this lecture for the purposes of this discussion is pp.385-95.

different political and non-political values relate to the political conception. These two stages seem to have been thought of as a kind of sequence. Hence, in the first instance, it is up to citizens individually as members of civil society to work out how their particular non-public viewpoints relate to the political conception: ‘the citizen accepts a political conception and fills out its justification by embedding it in some way into the citizen’s comprehensive doctrine as either true or reasonable, depending on what that doctrine allows.’⁷⁵ In this process citizens are to use the resources provided by their comprehensive doctrines, as the political conception does not offer any prescriptive way of accomplishing this. Then, in the third stage, we move further from the non-public point of view and look for a general public justification of justice. Here we deviate from the comprehensive doctrines that have supplied the ground for endorsing the political conception, and the focus is the mere fact of consensus (although of consensus obtained for the right reasons). Hence public justification occurs when all citizens have justified the political conception on the basis of their particular doctrines. This is an important step since ‘mutual accounting shapes the moral quality of the public culture of political society.’⁷⁶ Although comprehensive doctrines have an important role here, their particular contents do not play a normative role since citizens are asked to remain within the political domain. But no public justification can occur without the existence of an overlapping consensus. Rawls also makes two further, and crucial, connections. The first connection is between overlapping consensus and legitimacy. An overlapping consensus fulfils the conditions of legitimacy: ‘If we can make the case (for an overlapping consensus) then the conditions for (citizens) legitimately exercising coercive political power over

⁷⁵ Ibid., p.386.

⁷⁶ Ibid., p.387.

one another (...) are satisfied.’⁷⁷ The second connection is between overlapping consensus and stability. A reasonable consensus provides ‘the deepest and most reasonable basis of social unity available to us in a modern democracy.’⁷⁸ This results in stability for the right reasons. Hence we can see how justification, stability and legitimacy are all connected through the idea of overlapping consensus.

5. Conclusion.

This chapter has introduced the main elements of the project of political liberalism sequentially, as developed by Rawls in his successive writings. His initial concerns derived from his understanding of the role of political philosophy in a democratic society. In PnM, Rawls argues that political philosophy under democratic conditions should principally aim to uncover a publicly recognized basis for a conception of justice. Only a political (as opposed to a metaphysical) conception of justice can provide such a shared basis. From this it follows that justice as fairness must be understood as a political conception elaborated from the shared political ideas that belong to a specific tradition. Such shared ideas are loosely understood as the kernel of an overlapping consensus. The question of stability does not explicitly arise here, although there is a passing reference to the incompleteness of part III of TJ.

These themes are developed in IOC and connected together more closely. Rawls’s main aim here is to show that in a constitutional democracy the main role of political philosophy is to elaborate a political conception of justice capable of

- 1) providing a public basis of justification of its political and social institutions, and
- 2) guaranteeing that society’s long-term stability. Here Rawls explicitly acknowledges that it is the fact of pluralism what leads him to conceive of political philosophy in

⁷⁷ Ibid., p.390.

⁷⁸ Ibid., 391.

this way. His working hypothesis is that a political conception of justice will secure the stability of a constitutional democracy since there is no other possible alternative given the fact of pluralism. These remarks form the background for the development of overlapping consensus itself and for an assessment of some possible objections to this idea.

Most of the remaining elements of political liberalism are introduced in DPOC, in particular, the two stages in the presentation of justice as fairness, the burdens of reason, and the notion of constitutional essentials. Here Rawls also defines the domain of the political and clarifies why justice as fairness is not applied moral philosophy. Perhaps one of the most important points refers to what Rawls regards as the most appropriate terms of justification of political power. A liberal conception needs to be expressly designed to gain the support of citizens who have different comprehensive conceptions of value. There is an implicit connection here between overlapping consensus and justification.

Finally, PL introduces three crucial aspects of the project: the liberal principle of legitimacy, the fact of reasonable (as opposed to simple) pluralism, and the idea of constitutional consensus. Also crucially, in the 1996 edition (in particular in Lecture IX) Rawls explicitly connects overlapping consensus with the question of justification.

Now that our exegesis has been completed, I turn in the next chapter to examine and assess some crucial objections to the idea of overlapping consensus.

CHAPTER 2

OVERLAPPING CONSENSUS: A CRITICAL ENQUIRY

The first chapter has set the ground for my present discussion. Here I consider five different critiques directed against the idea of overlapping consensus. The layout of this chapter is as follows: the first part is devoted to an exposition of these five sets of criticisms, while in the second part the aim is to assess them in a way that is as favourable as possible to Rawls's idea. I conclude by asking whether or not there are still some remaining problems that may weaken his project.

1. Critiques of Overlapping Consensus.

In discussing the possible problems for overlapping consensus here, I do not aim to reach a definite conclusion, at least, not yet. In a sense, this is only a preliminary enquiry into the idea, and it will be necessary to wait until our examination of the notion of stability for a complete assessment of overlapping consensus. I now present the five critiques in a certain order. They have been ordered according to how hard it is for Rawls to answer them. Thus I begin with the least damaging argument and conclude with the most fundamental objection.

a) overlapping consensus is unnecessary because TJ had already solved the problem of stability

The argument, put forward by Brian Barry,¹ runs as follows: in view of the kind of stability that Rawls aims to achieve, the free-standing stage in the presentation of justice as fairness is in itself sufficient for the attainment of stability of the required kind. Here I am not concerned with examining Barry's complete review of Rawlsian stability, but only with those aspects that I consider crucial in his assessment of overlapping consensus. Also, there may be more than one way in which we could regard overlapping consensus as unnecessary, but only one case will be considered here, namely, Barry's claim that the second stage is superfluous for Rawls's own purposes of stability.²

Barry mounts his critique as a defence of the old Rawls against the new Rawls, or, to be precise, as a defence of most of the old Rawls. In Barry's view, much of what is problematic in PL (the second stage) is already advanced in chapter 9 of TJ (the congruence argument).³ Hence, the final chapter of TJ is the source of, what he regards as, unnecessary problems since overlapping consensus is mainly a review of the question of stability understood in terms of congruence. Just as chapter 9 was not needed in TJ to solve stability, the second stage of PL is also redundant. Let us see how Barry arrives at this conclusion.

In both TJ and PL Rawls generally regards stability as the existence of sufficient motivation on the part of citizens to follow the dictates of justice. Chapter 8 of TJ considered how in a well-ordered society citizens come to acquire the 'sense of justice', or the disposition to act in accordance with the dictates of justice. The

¹ This argument is provided in 'John Rawls and the Search for Stability', *Ethics*, 105 (1995), pp.874-915.

² *Ibid.*, p.891.

³ *Ibid.*, pp.883-87.

acquisition of a sense of justice leads citizens to see the demands of justice as overriding whichever other demands may conflict with it, including demands arising from their conceptions of the good. In Barry's view, this solves stability in TJ and nothing more is required. In fact, with regards to chapter 9, he recommends that we 'follow the course followed virtually unanimously by commentators on *A Theory of Justice* and forget about it.'⁴

According to Barry, Rawls is directed to the congruence argument by his rejection of the rational intuitionist 'doctrine of the pure conscientious act', or of the view that people are moved to act by their desire to do what is right simply because it is right and just, without taking into account other moral considerations.⁵ These other moral considerations and desires, i.e. the pursuit of human happiness, of equality and so on, are not motivations for the right itself. Rawls holds that this motivational condition, devoid of meaningful reasons, 'resembles a preference for tea rather than coffee.'⁶ Now Barry interprets this as Rawls's rejection of the view that acting purely out of a sense of duty is indeed rational, and he cites Scanlonian motivation (that the moral motive is the desire to act according to rules that could not reasonably be rejected by others similarly motivated) as a response to that rejection.

It is not the place here to either endorse or refute Barry's understanding of the rejection by Rawls of the doctrine of the pure conscientious act; what is of interest here is the conclusion that follows from this understanding. The conclusion is that in spite of Rawls's endorsement of the priority of the right over the good in both TJ and PL, he thinks that we need a further account of how people actually give priority to the right. The rejection of the doctrine of the pure conscientious act leads Rawls to adopt in chapter 9 of TJ an alternative doctrine, 'the ancient doctrine that no act can

⁴ Ibid., p.915n.

⁵ For Rawls's rejection of the doctrine of the pure conscientious act, see TJ, pp.477-479.

⁶ Ibid., p.478.

be regarded as rational unless it is for the good of the agent to perform it.’⁷ This in turn commits Rawls to provide an account of the links between justice and the good. In TJ this account is given by the congruence argument, and this is substituted in PL by the second stage of overlapping consensus.

Thus we have seen the relationship between the two books in a different light than the one provided by Rawls himself, but, what is the argument for the sufficiency of the free-standing stage and, hence, against the necessity of the second stage? The main point seems to be that the motivation for finding a shared basis of agreement on mutually reasonable terms is already present in the first-stage. We can see this in Rawls’s discussion of the ‘four facts’ of political sociology and human psychology that serve as the background for the free-standing stage in DPOC. These four facts together with the burdens of judgement underpin the condition that ‘people seeking agreement on reasonable terms with others must concede that the pursuit of their comprehensive views must be constrained by the limits laid down by Rawlsian justice.’⁸ Beyond this, Rawls only needs to provide an account of the moral motive (the basis of moral motivation), and he does this in Lecture II, just as in TJ he discussed the process of citizens’ acquisition of the sense of justice in chapter 8.

That Rawls regards the principles of right as the limit within which conceptions of the good are to be permitted can be seen in his discussion of the appropriate responses to someone who affirms the doctrine of *Extra ecclesiam nulla salus*. Leaving exegetical problems aside, there is, according to Barry, at least one version of this discussion that supports his point.⁹ This version suggests that political liberalism must abstain from making judgements about the truth or falsity of that

⁷ ‘The Search for Stability’, p.885.

⁸ Ibid., p.895.

⁹ Ibid., pp.898-901. This version is given in DPOC, p.483. Barry also discusses the other version given in IOC, p.435, which suggests that religious views that have illiberal implications are simply mistaken.

doctrine, but must affirm that it is unreasonable to use political power to endorse it as a basis for justice. This amounts to saying that what counts for endorsement (of the political conception) is not the content of doctrines, but their reasonableness, or, more to the point, the reasonableness of doctrine holders. However, the conditions of reasonableness are defined by the five facts of political sociology and human psychology, and these have already been built into the first stage. The first stage is therefore sufficient for stability. This can be seen even more clearly from Rawls's particular assessment of the links between the political conception and comprehensive doctrines. There he tries to do a balancing act by taking from each doctrine the version that may be more likely to support justice as fairness. However, if only the versions that are compatible with the political doctrine are selected, it is difficult to see why the second stage is required as the people holding those compatible doctrines' sub-sets would have already accepted justice as fairness in the free-standing stage¹⁰. An overlapping consensus is thus not required for Rawlsian stability.

b) overlapping consensus rests on a series of empirical inadequacies

Here I consider an argument made by Klosko who claims that there is crucial empirical evidence that demonstrates that Rawls's highest form of consensus is unattainable.¹¹ In order for an overlapping consensus to obtain, its content (justice as fairness) needs to be supported by most citizens. To avoid controversy, and given the conditions of reasonable pluralism, Rawls bases the content of the consensus on

¹⁰ A similar point is made by Mulhall and Swift when they suggest that 'the conditions that a conception must satisfy in order to be provisionally on hand after the first stage seem identical to the conditions that it must satisfy in order to be stable in the right way' and therefore 'his repeated urging of the two-stage presentation may seem less than helpful'; Mulhall, S. and Swift, A., *Liberals and Communitarians*, 2nd edition (Oxford, Blackwell, 1996), p.186.

¹¹ See 'Rawls's "Political" Philosophy and American Democracy', *American Political Science Review*, 87 (1993), pp.348-59, 'Rawls's argument from Political Stability', *Columbia Law Review*, 94 (1994), pp.1882-1897 and 'Political Constructivism in Rawls's *Political Liberalism*', *American Political Science Review*, 91 (1997), pp.635-46.

intuitive ideas that belong to the shared public culture of constitutional democracies. But Rawls provides a very specific understanding of the intuitive ideas running through justice as fairness and the two principles of justice. He thus makes a crucial factual claim: that the content of the consensus is already (at least implicitly) shared by citizens, even though citizens themselves may be unaware of this. The obvious move for Rawls would have been to test these empirical assumptions. His only attempt to do this is unsatisfactory as his investigation is restricted to the views of the Supreme Court regarding the right of free speech, and does not extend to the general views of the citizen body.¹²

The need for the empirical argument to work is derived from a philosophical point within Rawls's theory. Given his insistence on the burdens of judgement it is all the more pressing that he can show that citizens do in fact affirm his chosen set of values, otherwise they could be subject to the same degree of contestation as comprehensive doctrines.¹³ At this point Klosko draws on important social science empirical research that contains quantitative evidence that suggests that the principle of free speech (one of Rawls's basic liberties) is not as widely shared by citizens in some advanced democracies as Rawls seems to have thought.¹⁴ A series of surveys carried out between 1954 and, perhaps more relevantly, 1987 seem to contradict Rawls's optimism with regards to the existing level of consensus on basic political liberties. For instance, in the latter surveys conducted in Britain and America (1986 and 1987 respectively) just over a quarter of those interviewed were in favour of allowing members of groups they did not like to run for public office, and only half were in favour of allowing them to make a public speech. Generally, four political

¹² 'Rawls's "Political Philosophy" and American Democracy', p.351. (For Rawls's discussion of this point see PL, pp.343-44).

¹³ Ibid., p.352.

¹⁴ Ibid., pp.352-54.

liberties were surveyed (free speech, the right to demonstrate, the right to run for office and the right to associate), and these empirical studies consistently showed that the basic liberties are not as categorically affirmed as Rawls suggests. The upshot of this research is that an overlapping consensus is likely to fail given the *existing* degree of disagreement with regards to the basic liberties.

However, empirical research also shows a high degree of support for democratic procedures and broad democratic principles. A main problem seems to be that whilst most people would agree on the principle of free speech in the abstract, when confronted with particular proposals, many get 'cold feet'. For Klosko this suggests that citizens weigh 'strong' democratic rights against other values and considerations and citizens may prefer, for instance, to trade-off the right to free speech for the maintenance of public order; in this process, citizens may also take into account consequentialist considerations.

The problem for Rawls is that although some form of democratic consensus may be attainable and, therefore, may afford stability, democratic stability must not depend on a consensus (like overlapping consensus) that has, in Klosko's view, little chance of success. This evidence also shows that consensus on fairly demanding principles like Rawls's principles in fact amounts to endorsing a comprehensive view. The alternative would be a consensus on much thinner principles, a consensus that centres on citizens' support for the political system rather than on categorical moral principles.¹⁵ But this empirical problem does not only affect the second stage of consensus, but also the first stage. If the aim is consensus, he needs to ensure that the principles constructed in the free-standing stage do fit society's plurality of views,

¹⁵ The method that he favours is one of 'convergence', 'Political Constructivism in Rawls's *Political Liberalism*', p.638.

after all, Rawls is explicit about his aim of avoiding unnecessary controversies.¹⁶ This alternative type of consensus is within the reach of Rawls's theory, although it is not his preferred option. Klosko, following Baier, argues that a constitutional consensus, understood principally as an agreement on broad principles and democratic procedures, already exists, at least in the US and in other advanced democracies.¹⁷ A constitutional consensus has the advantage of not including matters of economic distribution as such highly controversial matters are not likely to generate widespread agreement.¹⁸ Moreover, in a constitutional consensus the content of basic rights is less defined and therefore, the conception of such rights is less categorical. Klosko also presents some empirical evidence that may suggest that citizens are generally inclined towards this weaker form of agreement. A series of surveys tried to gather data measuring the degree of support of American citizens towards their political system. The findings were generally indicative of the strong support given to the American political system. The different surveys consistently showed that over 80% of respondents were supportive of the different democratic institutions and of the system as a whole.¹⁹

In Klosko's view Rawls's insistence on overlapping consensus does come at a cost. Rawls is committed to abstain from comprehensive doctrines at the free-standing stage, however, he does not look at the political culture to see which ideas are actually

¹⁶ 'If we are less sanguine about this problem -Klosko states-, we can see that Rawls glosses over serious difficulties, especially the need to take appropriate steps in the first stage of construction to make sure the resultant principles will fit with society's view', *Ibid.*, p.638. My aim here is, of course, narrower than Klosko's as his main intention is to show the weaknesses of political constructivism.

¹⁷ *Ibid.*, p.639. See also Baier, K., 'Justice and the Aims of Political Philosophy', *Ethics*, 99 (1989), pp.771-90.

¹⁸ For instance, David Miller has shown that the difference principle does not command strong support, 'Distributive Justice: What the people think', *Ethics*, 102 (1992), pp.555-93.

¹⁹ 'Political Constructivism in Rawls's *Political Liberalism*', pp.639-40. The results of research undertaken in 1992 show, for instance, that Congress was generally approved by 88% of those surveyed, the Supreme Court by 94% and the Presidency by 96%, with only 8% disapproving of the constitutional structure of the US government. He terms this 'diffuse political support', see *Ibid.*, p.639 and 'Rawls's "Political" Philosophy and American Democracy', p.356.

held by citizens. This is also problematic for his conception of the person. Again, quantitative data indicates that over a third of Americans would reject Rawls's idea of the person on religious grounds.²⁰ Also, his exclusion of fundamentalists on the grounds of their unreasonableness pose problems for stability as this would imply that about 20% of American citizens are excluded.²¹ Hence empirical evidence seems to suggest that overlapping consensus is either unfeasible, or cannot afford stability without costs. A constitutional consensus may seem, from the same empirical point of view, more likely to afford stability.

c) overlapping consensus pre-empts the political

This critique focuses on the potential of Rawls's theory to engage in political debates, and, by extension, on the democratic credentials of Rawlsian liberalism.²² Two main sets of claims are directed against the second stage of justice as fairness on this ground: 1) an overlapping consensus excludes people's non-political ideas, and 2) it discourages radical social criticism. I shall consider each set of claims in turn.

We start with the first claim. A consensus that tries to bypass citizens' most deeply held values and ideas restricts the possibility of negotiation and compromise that form the daily stuff of democratic politics. The demand that citizens ignore the rest of their values when discussing political matters does in effect pre-empt the political domain.²³ Not only are values scrutinized before accessing any kind of public

²⁰ Ibid., pp.640-642.

²¹ Ibid., p.641.

²² Bellamy, R., *Liberalism and Pluralism* (London: Routledge, 1999), esp. chapter 2: 'Trimming values: Rawls and the constitutional avoidance of politics, pp.42-66, Moller Okin, S., 'Political Liberalism, Justice and Gender', *Ethics*, 105 (1994), pp.23-43, and Exdell, J., 'Feminism, Fundamentalism and Liberal Legitimacy', *Canadian Journal of Philosophy*, 24 (1994), pp.441-64.

²³ Bellamy compares Rawls's writings with Hayek's constitutional proposals. In both cases, Bellamy maintains, the result is 'a constitutional restriction of politics that severely circumscribes both the sphere within which the state may legitimately exercise its coercive power, and the kinds of considerations voters can invoke and legislators ought to take into account when making policy', 'Trimming values', p.42.

discourse, very relevant political questions (like the constitutional essentials) are also defined prior to citizens entering the political realm.²⁴ This, instead of leading towards stability, may result in a type of politics that takes citizens further away from any kind of active engagement.²⁵ Furthermore, it may seem that the exclusion of non-political values in practice endorses a view that supersedes the limits of political liberalism. The project of overlapping consensus thus appears self-defeating. Designed to secure stability in the face of disagreements, it cannot fulfil its role unless the very conditions for adopting it in the first place (the rejection of metaphysics) are in fact relaxed.²⁶

Hence, according to Rawls, under overlapping consensus stability is secured by the separation between the political and the non-political domains and by citizens' reliance on political values alone. This is a consequence of the fact of pluralism and of the other related facts. The problem here is that Rawls sees his solution as the only workable alternative (in view of these facts), an alternative adopted almost of logical necessity. There is a kind of categorical connection between the fact of pluralism and overlapping consensus. This is particularly evident in the case of the constitutional essentials. These must be fully elaborated and secured independently of people's general moral views. Because of this, Rawls presents the constitution as something already pre-defined. The foundations of a liberal democratic constitution rest on an evolutionary historical account of democratic development defined by a particular experience (the Anglo-American experience). As Bellamy suggests, 'Rawls seems to offer a Panglossian view of the liberal constitution –as though it could be founded on no lasting wrong, and (...), as though no regress from it were possible.'²⁷

²⁴ Ibid., pp.48,53 and 57.

²⁵ Ibid., p.43.

²⁶ Ibid., p.55.

²⁷ Ibid., p.50.

We know, however, that such a key constitutional essential as basic liberty is not beyond controversy, in particular with regards to its interpretation. Conflicts within the basic liberties are not easily resolvable by appealing only to political values.²⁸ We saw in the previous section how some citizens are willing to trade liberty for the sake of peaceful coexistence. Deciding on matters concerning the basic liberties may require some reliance on citizens' larger moral perspectives. If we look at the types of issues that legislative bodies are confronted with on a daily basis, we can see that debates around many of these issues are in actual fact debates about the interpretation and application of a basic liberty. Not all matters involving the basic liberties can be regarded as constitutional essentials, as Rawls suggests. Whether it be ID Cards, stem cell research, the treatment of suspected terrorists, or pornography, political debate needs to confront disagreement -even on questions involving the basic liberties- rather than to brush it aside. This would entail the removal of the basic liberties from the 'metapolitical level of overlapping consensus'²⁹ and their firm placement in the realm of everyday politics.

So far we could just say that overlapping consensus rests on a mistaken view of how politics functions. However, the problem turns out to be more substantive than this given Rawls's insistence on the burdens of judgement. There is nothing in the definition and layout of the burdens of judgement that prevents their application to Rawls's interpretation of the basic liberties and, more generally, to the constitutional essentials. If we take the issue of physician-assisted suicide, something that Rawls regards in certain cases as a constitutional essential, we can see how people might

²⁸ Bellamy illustrates this with the example of the Rushdie affair. If we understand the conflict mainly as a 'clash of truths', 'Rawls either has to argue that the political liberties have some higher worth, or concede that those taking offence have valid interests at stake. Either way, he will be drawn into arguments concerning the public good and perfectionist values, even if it is to justify the virtues of political debate', *Ibid.*, p.57.

²⁹ *Ibid.*, p.53.

disagree about the relative weigh of the different philosophical, ethical and religious considerations.³⁰ Moreover, they might disagree as to whether the Supreme Court (or any other court) should decide on the particular cases of physician-assisted suicide that may be constitutionally protected. All this appears to show a lack of trust in the *demos* which is, after all, the basis of any liberal democratic constitution. In the end, and this would appear to be fatal to the project of political liberalism, Rawls needs to resort to metaphysics in order to justify the exclusion of citizens' conceptions of the good from the realm of politics. The Rawlsian political self is a metaphysically loaded character that needs to be able to distinguish between: 1)different kinds of political values, 2)political and non-political values, and 3)constitutional and legislative issues. Not only that, this political self is required to apply an appropriate set of values to the appropriate set of questions.³¹ An overlapping consensus therefore leaves very little room for politics.

We now come to our second claim, namely, that overlapping consensus discourages radical social and political criticism. Whereas the conclusion that we can draw after outlining our first claim is that Rawls's second stage poses too many restrictions on the democratic functioning of a polity, this second charge manifests an inverse objection, namely, that overlapping consensus gives in too much to the fact of pluralism. Because the fact of pluralism underwrites the liberal principle of legitimacy according to which constitutional principles must be generally acceptable, Rawls's theory is unable to deal with radical critiques, such as the feminist critique of the family and of the social institutionalisation of gender, or critiques of the role of the market economy in promoting political equality. Since such critiques are inspired by

³⁰ See the 1998 'Commonweal interview with John Rawls' in *Collected Papers*, edited by Samuel Freeman (Cambridge, Mass.: Harvard University Press, 1999), pp.618-19.

³¹ 'Trimming Values', pp.57-63.

grave concerns with issues of (in)justice, it is all the more pressing that a theory of justice, such as Rawls's, is able to at least assess their relative merits and incorporate into his own project whichever aspects are essential to a just society. But this objection goes further than that. Some commentators have argued that in seeking a consensus, 'Rawls's theory of justice lends itself to the goal of accommodation with religious conservatives by limiting state power to effect a feminist reform of domestic gender roles', and therefore 'to achieve justice for women, liberals must abandon the search for a legitimating consensus and take sides in what may be the most impassioned cultural and political struggle of our time.'³² Ultimately there is a cultural clash between conservatives and feminists that can be easily translated into a series of fundamental political disagreements; this results in it being impossible for both positions to coexist within an overlapping consensus. As Moller Okin has suggested, whereas in TJ it was possible to apply principles of justice to the family via the original position, in PL this task is undermined by some of the central elements of Rawls's revised framework, in particular by his relegation of the family to the non-political sphere.³³

Strictly speaking the family belongs to society's basic structure and is therefore regulated by a political conception of justice in so far as it performs the socially necessary function of ensuring society's long-term viability through the reproduction and nurture of future citizens. It is in this restrictive sense of making sure that future citizens have adequate opportunities to develop that principles of justice do also apply to the family. However, a political conception of justice is not designed in the first place to apply to the internal affairs of the different social institutions including the family. Hence Rawls insists here in the public and private

³² 'Feminism, Fundamentalism, and Liberal Legitimacy', p.442.

³³ 'Political Liberalism, Justice and Gender', pp.24-25.

dichotomy. As citizens, all adult family members enjoy equal rights and opportunities, and all children enjoy equal protection by the state. The problem is that some of these legal rights may become worthless since they cannot be applied to the internal affairs of families where many of the sources of injustice originate. This would seem a concession to those ways of life that are on the margins of democracy and therefore an attempt to win them around and ensure their acceptance, or at least their tolerance, of liberal democracy. A more comprehensive gender equality is thus sacrificed for the sake of a wider consensus. Rawls admits this quite candidly when he says that 'a liberal conception of justice may have to allow for some traditional gendered division of labour within families (...) provided it is fully voluntary and does not result from or lead to injustice.'³⁴ And, if there are doubts as to what being voluntary exactly means, he goes on to say that it 'means that it is adopted by people on the basis of their religion, which from a political point of view is voluntary and not because various other forms of discrimination elsewhere in the social system make it rational and less costly for husband and wife to follow a gendered division of labour in the family.'³⁵

A feminist critique would respond that although traditional roles may have been adopted more or less willingly, and not as a result of the use of coercion, they nonetheless would have been inculcated through a process of indoctrination. A society aiming to be just needs to at least examine the nature and result of such processes, and to ask questions about the limits of religious toleration particularly when the free exercise of religion conflicts with the equality of women. In the meantime, a theory of justice whose main goal is to obtain as wide a political consensus as possible will

³⁴ IPPR, p.599.

³⁵ Ibid., pp.599-600.

continue to exclude controversial matters by defining them as non-political. Such exclusion would call into question the very nature of a just society.

d)an overlapping consensus stops short of being a normative idea

In this section we are concerned with one aspect of Habermas's critique of Rawls. To see how this critique has been mounted, we should begin by recalling some of Rawls's own remarks on the relationship between overlapping consensus and stability. In PL stability has two functions; the first is compliance and the second is consensus (or compliance for the right reasons).³⁶ The first function of stability involves examining 'whether people who grow up under just institutions (...) acquire a normally sufficient sense of justice so that they generally comply with those institutions.'³⁷In this respect stability in PL is not dissimilar from stability in TJ. But what interests us here is the second function of stability (consensus) given that it departs in important ways from TJ. The suggestion here seems to be that general compliance is not sufficient because the fact of reasonable pluralism cuts across people's sense of justice. People's sense of justice needs to be integrated within their respective comprehensive doctrines (provided they are reasonable) for consensus to obtain.

Another important point stressed by Rawls is that, given that considerations of stability only shape his theory at the second stage, the content of justice as fairness is not influenced by any type of contingency regarding the share of power between the different comprehensive doctrines. The relevance of this second stage is that only by addressing each citizen's reason, can justice as fairness be considered a reasonable

³⁶ PL, pp.140-44.

³⁷ Ibid., p.142.

conception of justice.³⁸ Hence, if 1) the content of justice as fairness is worked out independently of whatever other general religious, philosophical and moral views citizens happen to have, and if 2) justice as fairness needs to elicit reasoned support from those general views, Rawls seems to be pointing to the development of a kind of justificatory attitude on the part of citizens. This becomes clearer when he explicitly concludes that only a theory endorsed in this way (by reference to each citizen's overall moral viewpoint) satisfies a liberal account of the legitimacy of political authority. An overlapping consensus therefore also uncovers a public basis of justification.³⁹

It is in this context that Habermas asks whether overlapping consensus is to be understood primarily as a stabilizing mechanism, or as a justificatory device.⁴⁰ As we have seen, there are grounds for suggesting that Rawls has some kind of justificatory role in mind beyond the attainment of mere compliance. But this is in itself problematic since given the very different nature of the two roles, it is difficult to see how overlapping consensus can solve them both.

From Habermas's point of view, Rawls appears to have mistakenly assumed that his initial solution to the problem of stability (the successful acquisition of the sense of justice by the members of a well-ordered society) can be extended to include overlapping consensus. There is a significant difference between the test of stability (understood in terms of compliance) and the test of consensus in Rawls's theory. Whereas the test of stability is a step taken within the theory itself by looking at how people who live under just institutions come to acquire the required sense of justice so that they comply with those just institutions, the test of consensus is external to justice as fairness as it focuses on the possible links between justice as fairness and

³⁸ Ibid., p.143.

³⁹ Ibid., p.144.

⁴⁰ 'Reconciliation Through the Public Use of Reason', p.119.

comprehensive doctrines. To put it in Habermas's terms, whereas the test of acceptance undertaken in TJ was internal to the theory, the test of acceptability (overlapping consensus) goes beyond the theory itself.⁴¹ It is a move external to justice as fairness, and Rawls assumes that both tests are of a similar nature.

If the function of overlapping consensus is mainly to solve stability, then the question of the theory's justification must have already been settled. As we know, a self-referential type of justification has already been established. But then Rawls still needs to settle the kind of external justification that we suggested above and that he also seems to have in mind. But now the problem is this. Rawls does not really pose any real constraints within overlapping consensus; he wishes to leave the result of the exercise open so as to not overburden it for the sake of stability. But then it is difficult to see what the theory gains with this exercise. If stability has already been settled by citizens acquiring the sense of justice, and if substantial epistemological connections between justice as fairness and comprehensive doctrines are ruled out, what precisely is the role Rawls gives to overlapping consensus? Habermas's suggestion is that if a merely pragmatic role is rejected, then Rawls needs to show that the validity of justice as fairness does not depend on particular comprehensive doctrines. He needs to provide an epistemic link, through the category of 'reasonableness', between normative validity and neutrality towards conflicting comprehensive doctrines.⁴² Only in this way would the account cease to be merely pragmatic. But this move is disallowed by Rawls's assignation of the category of 'truth' to comprehensive doctrines. This is, in Habermas's view, deeply problematic.⁴³ Political conceptions do not respond to truth, but to reasonableness. Reasonableness may well be understood in terms of normative validity. Hence if comprehensive doctrines are true (or false),

⁴¹ Ibid., p.120.

⁴² Ibid., p.122.

⁴³ Ibid., p.124.

justice as fairness could only be true if it was derived from a true conception. However, justice as fairness aims to be neutral towards comprehensive doctrines and therefore its reasonableness cannot be established in this way. Hence we must understand reasonableness in a much weaker sense. A reasonable conception of justice is such that tolerates comprehensive doctrines that are not unreasonable. Understood in this way, an overlapping consensus fails to afford adequate normative credentials.

e)an overlapping consensus is utopian⁴⁴

We turn now to the final objection to the idea of overlapping consensus, an objection originally suggested by Baier and acknowledged by Rawls himself.⁴⁵ In a key 1989 article Baier asks whether Rawls has provided an empirical account of how a conception of justice can become the focus of an overlapping consensus, or whether he intends to leave this question open. It seems that Rawls faces a dilemma: on the one hand, he claims that his conception could be accepted by a substantial number of citizens because its foundations lay in the public political culture that we all implicitly share. On the other hand, however, Rawls does not work out the details of how particular comprehensive doctrines may come to accept the conception of justice as he regards this as an unnecessary step. The important thing is the fact of consensus itself; the fact of consensus amongst different moral positions. But, if his concern is stability (a practical concern after all), doesn't he need to define the process whereby his political conception is accepted by most citizens, and, moreover, to show that it provides a sufficient basis for stability? As Baier suggests, 'if he is satisfied with a

⁴⁴ Just to clarify the use of terminology here: references to the problem of utopianism must be understood in terms of whether justice as fairness can be considered 'unrealistically utopian'. I think that this is the way in which we need to understand Rawls's discussion in PL, pp.158-64.

⁴⁵ 'Justice and the Aims of Political Philosophy'.

merely imaginatively realizable or foreseeable consensus (...) he [may be] open to the charge of utopianism...'⁴⁶

Perhaps the most complete definition of the problem of utopianism has been provided by Nagel.⁴⁷ A specific project of social and political change may be considered utopian (in the 'pejorative' sense of being unrealistically utopian) if it cannot generate its own support. What is needed here is both moral and political viability and this depends on the extent to which moral justification, individual motivation and political institutions all work in tandem to support such project of social change. To ignore moral justification would be to completely give in to the demands of human nature, and ignoring individual motivation (people as they are) runs the risk of sidestepping the complexity of our predicament. The role of institutional arrangements is important to balance possible unrestrained human motives, but also to constrain unrealisable projects. What this entails is that we must seek for dual justifications. In political theory, justification must firstly be addressed to the impartial standpoint, and secondly, to each of us, as individuals, but also as members of a specific institutional setting. In Nagel's view, this second task is necessary if the charge of utopianism is to be avoided.

These ideas are echoed by Rawls's own acknowledgement of the problem. I have already explained the solution that he proposes and the role of constitutional consensus within it. Also Baier has suggested that, by making stability dependent on an overlapping consensus, Rawls might have set the stakes too high, and a constitutional consensus, of the kind that already exists in most advanced democracies, might be a more realistic expectation from the point of view of

⁴⁶ Ibid., p.783.

⁴⁷ Nagel, T., *Equality and Partiality* (New York: Oxford University Press, 1991), esp. chapter 3 'The Problem of Utopianism'.

stability.⁴⁸ Such a consensus, less specific and less intense, would be, as a result, a more extensive, and therefore inclusive, consensus. Also a constitutional consensus would avoid potential conflicts with religion and metaphysics as it does not depend on other doctrines for support. A constitutional consensus would seem, all things considered, a more adequate way of achieving stability.

We know that Rawls takes the idea of constitutional consensus on board, although not as a way of dealing with stability, but with utopianism. A constitutional consensus cannot solve stability because it lacks the necessary moral credentials that overlapping consensus has. Only a consensus with appropriate moral foundations can do the job of securing long-term stability. Hence a constitutional consensus merely functions as an explanatory idea, with minimal normative connotations. Rawls thinks, however, that by telling an evolutionary story of how a *modus vivendi* becomes an overlapping consensus, he has provided a satisfactory answer to the charge of utopianism⁴⁹. An overlapping consensus is utopian if such a consensus proves to be either unfeasible or unstable. A *modus vivendi* is not a consensus in the Rawlsian sense, but if it can be shown how a *modus vivendi* turns into a constitutional consensus over time, then the charge of unfeasibility could be escaped. Similarly, if it is possible to show how this more feasible constitutional consensus finally develops into an overlapping consensus, the stability problem will have been settled, and the charge of utopianism rebutted. The key factor in this process is the partial comprehensiveness of most doctrines so that most people are able to accept the

⁴⁸ 'Justice and the Aims of Political Philosophy', p.775.

⁴⁹ The point just made about the minimal normative weight of constitutional consensus needs some refinement. This is because the related concept of constitutional essentials does seem to play a key normative role in Rawls's theory. However, the idea of constitutional essentials is not discussed in relation to the problem of utopianism (or, to stability for that matter), and is instead related to the notions of public reason, political legitimacy and public justification. See PL, pp.227-30. Chapters 5 and 6 will be devoted to some of these questions.

political conception for its own sake.⁵⁰ This in turn leads to a consensus on the constitutional arrangements that people have willingly accepted once they have experienced their benefits. This starts a process of political cooperation that culminates with an overlapping consensus. For Rawls this sketch is in itself sufficient to answer the charge of utopianism.

Although Rawls has always remained quite optimistic about the prospects for overlapping consensus on his two principles of justice, he nonetheless admits that 'if the liberal conceptions correctly framed from fundamental ideas of a democratic public culture are supported by and encourage deeply conflicting political and economic interests, and if there be no way of designing a constitutional regime so as to overcome that, a full overlapping consensus cannot (...) be achieved.'⁵¹ On the other hand, Rawls affirms that the main reason for the introduction of overlapping consensus is to make the idea of the well-ordered society more *realistic* given the social and historical features of advanced democratic societies.⁵² So how can a theory be reformed in such a way as to become more realistic, and then claim that these reforms make the theory utopian? This appears to some extent contradictory.

This pessimism about the prospects for overlapping consensus is shared, for instance, by Scheffler, who suggests that Rawls's model conception of comprehensive doctrines may be too narrow, and may be only a partial reflection of the extent of value pluralism within liberal societies. But, even if this model is expanded, it appears unlikely that utilitarianism would figure in the consensus given that the main arguments for the two principles of justice are elaborated as an alternative to utilitarianism itself.⁵³ Moreover, as Rex Martin has suggested, utilitarianism,

⁵⁰ Ibid., pp.158-164.

⁵¹ Ibid., p.168.

⁵² Ibid., p.19 and JFR, p.32.

⁵³ Scheffler, S., 'The Appeal of Political Liberalism', *Ethics*, 105 (1994), pp.4-22. See p.9.

including Mill's version, cannot guarantee their support for the priority of constitutional rights over considerations of aggregative welfare.⁵⁴ Brian Barry also claims that whereas some version of utilitarianism may endorse Rawls's first principle, other versions may even reject liberty and democracy.⁵⁵

So far the problem has been whether, and to what extent, the content of the political conception can be endorsed by different comprehensive doctrines. But there is also an added difficulty. Some doctrines may endorse the two principles of justice, but may be reluctant to accept the political status of Rawls's conception. A consensus that requires the acceptance of a fairly substantial metathesis over and above the content of the agreement would seem to be unnecessarily exclusive.⁵⁶ Wenar, focusing on the particular case of the major religions, powerfully argues that whereas most religions could endorse Rawls's two principles and, also, his conception of the person, Rawls's emphasis on a heavily loaded conception of reasonableness prevents major religions from being part of an overlapping consensus.⁵⁷ Wenar signals as problematic elements the idea of the burdens of judgement, the account of reasonable moral psychology and Rawls's views on constructivism and objectivity. These three elements make the theory a partially comprehensive one, and this makes it unsuitable as the focus of an overlapping consensus. Brian Barry goes even further suggesting that 'there is no reason in general for expecting religions, even if they accept the "doctrine of free faith", to endorse the equal civil and political rights that make up Rawls's first principle of justice.⁵⁸ As far as the second principle of justice is

⁵⁴ Martin, R., 'Rawls's New Theory of Justice', *Chicago-Kent Law Review*, 69 (1994), pp.737-61. See pp.757-61.

⁵⁵ 'The Search for Stability', p.907.

⁵⁶ As Scheffler has suggested, this is not only a problem for the feasibility of the required consensus, but 'any requirement that the participants in an overlapping consensus must view the conception of justice as political would appear incongruous with the motivation for introducing the idea of such a consensus in the first place'. 'The Appeal of Political Liberalism', p.14.

⁵⁷ Wenar, L., 'Political Liberalism: An Internal Critique', *Ethics*, 106 (1995), pp.32-62.

⁵⁸ 'The Search for Stability', pp.909-10.

concerned, it is significant that Rawls does not even attempt to show that religious views incorporating an “account of free faith” must underwrite it’.⁵⁹ Therefore, all things considered, an overlapping consensus appears too utopian. Let us now see whether Rawls’s theory can defend itself against this and the other four objections.

2.Overlapping consensus: A critical assessment.

Here I shall consider each objection in turn. My aim is to defend as far as possible Rawls’s second stage of his theory. I anticipate that there may be some important difficulties for the idea of overlapping consensus, however, a full assessment will take place once the companion idea of stability is examined in the next chapter.

a)the second stage of justice as fairness is superfluous in Rawls’s theory

I now try to present a reply to the view that the second stage is unnecessary since the motivation for a shared basis of agreement was already present in the first stage, and therefore, those who are reasonable would have already endorsed justice as fairness. I will assume for the sake of this discussion that overlapping consensus is, as Barry claims, a substitution for the congruence argument given in TJ.

In order to respond to this objection we need to recall Rawls’s comments regarding the nature of political philosophy.⁶⁰ In a democratic setting political philosophy fulfils a number of roles. The first role is the practical one of avoiding controversy and of searching for an adequate basis of agreement. When this is not possible, political philosophy should at least specify the terms of some kind of political cooperation, no matter how narrow those terms may be. The concern is therefore to surpass conflict. As we know, Rawls is primarily concerned with that

⁵⁹ Ibid.

⁶⁰ JFR, pp.1-5.

kind of conflict that is rooted in reason. This is the highest level of conflict, placed above conflicts of interests or even theoretical disagreements. This is a conflict about 'the different philosophical and moral doctrines that deal with how the competing claims of liberty and equality are to be understood, how they are to be ordered and weighed against each other, and how any particular way of ordering them is to be justified.'⁶¹ In Rawls's view this highest level of conflict in a sense defines the other more narrow types. The practical aim of finding a suitable basis for political agreement cannot be fulfilled by any one of those moral or philosophical theories, and Rawls's first stage is precisely designed to avoid the deepest roots of conflict and it must therefore be understood in the light of this practical aim.

The other role of political philosophy that interests us here is that of being a realistic utopia.⁶² Political philosophy needs to examine practical possibilities in order to account for people and societies as they are. The stakes are set high when thinking of the necessary requirements of a just and democratic society, however, we must account for the circumstances of justice for there is no point in imagining what Peter Laslett refers to as 'the world that we have lost.'⁶³ The circumstances of justice suggest that pluralism is an obstinate fact that is likely to persist under democratic conditions. We are therefore led to view the second stage of overlapping consensus as an attempt to show how ideal principles can become practical possibilities.

It might be said that these remarks do not change the nature of Barry's objection but, in fact, reaffirm it. For, if the practical aim of finding agreement has been present at the outset, this would greatly benefit the prospects for the adopted principles. But note, that there are two ways of fulfilling this practical aim. One is to

⁶¹ Ibid., p.2.

⁶² Ibid., p.4, and *The Law of Peoples* including the paper 'The Idea of Public Reason Revisited' (Cambridge, Mass.: Harvard University Press, 2000), 4, 5-6, 11-12.

⁶³ Laslett, P., *The world we have lost* (London: Methuen, 1965).

remove the possible obstacles on the way to it, and the other is to find a compromise or accommodation of existing positions. This is significant because if Rawls were to follow the second strategy of accommodation, then we could generally say that the second stage was redundant. In fact, we would not have a reason for dividing the theory into two stages. The strategy of avoidance, however, works in a qualitatively different way, firstly by removing the deepest roots of controversy so as to find an appropriate basis of agreement, and secondly, by bringing them back into play to test whether such an agreement is practically achievable.

Still it can be suggested, as Barry does, that discussion of the actual possibilities of justice as fairness very much depends on the theoretical assumptions present at the first stage. In other words, only those who are either Kantian or political liberals will be able to endorse a conception of justice such as Rawls's. Of course, one could be a political liberal and also affirm a general philosophical, moral or religious doctrine. This is unproblematic for Rawls (for endorsement of his principles of justice), but problematic for Barry as it suggests a kind of circularity within Rawls's overall framework. However, and this is all we need to say for now, there may be other types of disagreement beyond that between comprehensive doctrines. In other words, the first stage has been designed to surpass one kind of conflict but, in the second stage, other kinds of disagreement may come into play. And, we may suggest here two examples: disagreements generated by conflicts of interests, and relevant theoretical disagreements. These examples show another way of looking at the relationship between the two stages that suggests their relative inter-dependence, but not their circularity. Hence we may conclude that the second stage is necessary for Rawls's overall aim, otherwise his theory could not perform the tasks that it was set out to perform, that of showing that political liberalism is a realistic utopia.

What I am suggesting is that the second stage does not get exhausted with the idea of overlapping consensus. Even if we claim that overlapping consensus is unnecessary for the required kind of stability, we can still retain the aim of showing how justice as fairness can be politically justified. Furthermore, the second stage can be interpreted as a means of accounting for the democratic possibilities of justice as fairness. This is not only a legitimate purpose, but also a necessary one.

b) the empirical basis of Rawls's second stage

According to Klosko, a major fault line in PL is that the principles elaborated at the first stage do not take into account the actual views generally held in society. If Rawls were really concerned about consensus building, a method of convergence would seem more appropriate to obtain stability. Convergence would entail a reversal in the order of the two stages. A method of convergence would firstly try to highlight substantive points of agreement amongst comprehensive doctrines with the aim of finding, in the second stage, the most sound and robust normative principles that existing agreement allows. As we saw above, Klosko thinks that substantive agreement does exist and that Rawls, through the idea of constitutional consensus, has the theoretical resources to uncover its basis. I now present two replies to this objection.

The first reply is quite straightforward. Klosko's argument rests on a misunderstanding of the (Rawlsian) notion of constitutional consensus. He prioritises constitutional consensus because it only covers democratic procedures and decision-making mechanisms. But a constitutional consensus covers far more than mere democratic procedures. It covers the 'equal basic rights and liberties of citizenship

that legislative majorities are to respect.’⁶⁴ These include the right to vote (which Klosko would consider relatively uncontroversial), but also the right to free speech (which, according to this objection, is deeply problematic). These *essentials* are non-negotiable within a constitutional consensus and, therefore, the idea of ‘diffuse support’, that Klosko uses to prioritise constitutional consensus over overlapping consensus, would lead towards a *modus vivendi* type of agreement –an agreement based on a convergence of interests.

The second reply is, I think, more fundamental. We established in our answer to the previous objection that Rawls does not work out the first stage with the aim of agreement in mind. The move to the political is just a way of avoiding unnecessary, but reasonable, conflict. His aim is, rather, to offer a plausible account of how the values of freedom and equality can be articulated within a more or less complete conception of justice. These values are shared, and Klosko acknowledges this, but their specific Rawlsian articulation might not be shared, and this is, I think, Klosko’s precise point. Freedom and equality are too abstract to claim for them a kind of non-contestable status. A democratic society contains within itself different ways of understanding these values and, more importantly, their normative implications. So the problem is not so much the selection of values, but their particular construction within Rawls’s theory –a construction with no empirical basis. But this rests on a misunderstanding of Rawls’s project. Firstly, we can say that just because a project seems unattainable (assuming for the sake of the argument that it is unattainable), it does not mean that we surrender altogether the task of trying to find principles that reflect *the most adequate understanding of freedom and equality all things considered*. Such a project would be political ‘in the wrong way’. After all, one of the

⁶⁴ PL, p.227.

roles of political philosophy is to *orient* citizens in their political life; it must aid citizens' understandings of their common institutions and practices. This is done 'by specifying principles to identify reasonable and rational ends of those various kinds (individual and associational, political and social), and by showing how those ends *can* cohere within a well-articulated conception of a just and reasonable society.'⁶⁵ *Can* is important here as Rawls does not claim that particular ends *will* cohere with his conception of justice as suggested in this objection⁶⁶. Hence, political philosophy in Rawls's view also performs a crucial pedagogical role. This objection therefore seems to rest on a misreading of the project of political liberalism. Ultimately, what makes a theory of justice realistically utopian is its ability to probe the limits of practical possibilities. And this is certainly a legitimate aim of political philosophy.

c) a consensus that bypasses the political

This objection contains two inverse claims. On the one hand, the overlapping consensus is too restrictive because it does not allow particular values and ideals to filter into the political. On the other hand, the overlapping consensus is too permissive. The degree of inclusiveness of such a consensus necessarily tames radical political criticism. In this section I aim to show that Rawls's theory is not vulnerable to these objections.

Firstly we need to try to elucidate the sense in which a type of consensus is both too exclusive and too inclusive. As far as these objections go, we can say that consensus is too exclusive because people who take part in it have to abandon their general moral views when thinking and acting politically. By contrast, overlapping consensus is too inclusive because, in view of the goal of stability, it has to cater for

⁶⁵ JFR, p.3.

⁶⁶ Klosko, on the other hand, claims that particular ends *should* cohere with normative principles if consensus is to be attained.

people of diverse, and often antagonistic, persuasions. As we saw in the first part of this chapter, the real danger here is that certain political goals, i.e. a feminist redefinition of domestic roles that supersedes traditional gender divisions, may be considered too controversial, particularly by holders of religious comprehensive doctrines and, therefore, basic issues of justice may have to be abandoned for the sake of a stability-seeking consensus. But, if people have to set aside their non-political values in any case (as the first part of this objection states), it will seem that this criticism is misplaced. Of course, it may well be the case that radical reforms cannot be put into practice for other reasons of a generally political kind. However, if people's non-political values are allowed to play a part, as the second claim suggests, the first part of this objection will not hold. It seems that, by contrast with Bellamy's interpretation, the two parts of the objection cannot hold simultaneously.

I would like to suggest a more helpful way of putting these points. The first part of the objection stems from a concern with personal liberty whereas the second derives from a concern with political equality or, to put it differently, the first part of the objection relates to people's sense of the good whereas the second part focuses on claims about justice. We need to recall here two of the implications of Rawls's conception of the person as free and equal. Firstly, citizens are free because they see themselves and one another as capable of having both a sense of justice and a conception of the good, and secondly, citizens are free to make claims on their institutions with a view to advance their particular conceptions of the good. People are thus seen as having both a political and a non-political identity. The duties and obligations derived from these two kinds of identities that together give a person's life its moral worth must be somehow reconciled.⁶⁷ So, if citizens are entitled to make

⁶⁷ PL, p.31.

claims on institutions based on their conceptions of the good, it is reasonable to think that general moral values and beliefs are going to access certain public discourses. If having a conception of the good is an important part of the moral significance attached to citizens' lives, public discourses that are obstructive of non-political values will not aid the proposed reconciliation and adjustment between people's public and non-public identities. Likewise, if citizens are regarded as capable of having a sense of justice, being self-authenticating sources of valid claims entitles them to articulate radical political proposals. Rawls does not contemplate restrictions on the content of political speech, even in the case of advocacy of revolutionary or seditious doctrines.⁶⁸

So where have the critics gone wrong? Or, is there a contradiction between different parts of Rawls's theory? Does the second stage of overlapping consensus modify some of the basic ideas mentioned above? It seems that the idea of the person as a self-authenticating source of valid claims does some crucial work within Rawls's theory and, therefore, any deviation from it must be carefully substantiated. So are there any grounds for thinking that this basic idea has been watered down at the second stage of the theory?

In thinking of an overlapping consensus, Rawls aims to show that his political ideal of free and equal persons is compatible with most people's general moral conceptions. The presumption is that it is, and thus the burden of proof falls on comprehensive doctrines. The intuition behind this is that there may be many competing metaphysical, moral and/or religious understandings of free and equal, but that most would accept a political ideal of free and equal as a pre-requisite for the exercise of their own particular (and more substantial) conceptions of free and equal.

⁶⁸ Ibid., p.336.

However, according to Bellamy, even this political conception is metaphysically loaded since citizens must be able to distinguish between the commonly shared political ideal and their diverse metaphysical or religious understandings of freedom and equality. But, whereas some distinction between political and non-political values may be required at some point, this does not occur at the level of theory formation since the political conception of the person serves mainly to emphasize the democratic character of Rawls's theory. As I said above, the burden of proof is on comprehensive doctrines; it is up to them to show how and why this political ideal is not acceptable to them. It is assumed that most comprehensive doctrines would accept it if only for instrumental reasons, i.e. as a guarantor of their own survival in a pluralist society. Hence there is no metaphysical view invoked at the first stage of the theory, and a very definite political (and democratic) ideal put into place. Being part of an overlapping consensus requires the prior acceptance of this basic political ideal that enables citizens to be self-authenticating sources of valid claims under conditions of equal footing. Furthermore, although some distinction between political and non-political values may be operative at some stage, this is not something that is peculiar to Rawls's theory, but to the way in which most democratic polities function. This requires that citizens articulate their claims with regards to justice in political terms. Citizens do not have to operate a personality split between their political and non-political identities. It seems quite likely that the particular claims advanced may still be furnished by citizens' conception of the good. Rawls's theory can accept this provided the claims are politically relevant. In the end Bellamy's objection is motivated by his view of democracy –a view that differs from that of Rawls. Bellamy's conception, where every political issue, including the constitutional essentials, is up for grabs and where majority rule is paramount, may be an attractive

one, but, according to Rawls, would not, amongst other things, afford the necessary stability to a polity since it will be very much dependent on the particular balance of forces operating at any given time. Thus Rawls, as Amy Gutmann has suggested, 'joins most democratic theorists in not valuing majority rule so highly as to place it above the protection of equal liberty or the realization of just legislation.'⁶⁹

Furthermore, Rawls now acknowledges the wide view of public political culture.⁷⁰ This implies that arguments based on comprehensive doctrines, both religious and non-religious, may have a place in the public deliberations of political matters. This more inclusive view is only restricted by the so-called *proviso*, which states that proper political reasons must be also supplied *ex post*. The *proviso* is necessary because of political liberalism's commitment to public justification. However, allowing citizens to appeal to their non-political values (or to comprehensive political values) is not based on a necessity to compromise on Rawls's part, but on the recognition that comprehensive doctrines themselves often provide the basis for citizens' allegiance to a democratic polity. Allowing arguments from comprehensive doctrines in the political forum not only strengthens the idea of public reason, but also our commitments as members of a democratic citizenry⁷¹.

To conclude this assessment, it remains to be shown that radical political criticism is possible within the terms of political liberalism. Recall that the objection focused on why a redefinition of gender roles was not included in the theory of justice, and the suggestion was that through the idea of overlapping consensus Rawls had tacitly given in to religious conservatives. Here we need to ask why a more robust treatment of gender equality was not included in PL, and also whether the theory has

⁶⁹ Gutmann, A., 'Rawls on the Relationship between Liberalism and Democracy', in Freeman, S. (ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), p.189.

⁷⁰ IPPR, pp.591-94.

⁷¹ I discuss public reason at some length in chapter 6. Here I am only advancing possible arguments to respond to Bellamy's objection.

the potential to actively advocate a redefinition of gender roles. This second aspect has already been answered above. If Rawls allows even revolutionary and seditious doctrines into the public realm, then it is difficult to see why he would not allow any kind of radical claim such as the one we are considering. With regards to the first question, we can say that Rawls's main concern is with justice, and gender equality is certainly a part of justice. However, Rawlsian justice contains two basic elements: personal liberty and political equality. The suggestion implied by the critics is that personal liberty seems to override any claims derived from a concern with political equality, and particularly with gender equality. But is this always and generally the case? Critics may have been led to this view because Rawls has not included any specific issue regarding gender identity into the list of constitutional essentials, and, since this idea is of utmost importance within PL, gender equality would appear to have lost some ground with respect to TJ. Here we are beginning to touch on substantive issues that will be addressed later on in this thesis. Since at present we are concerned with the effects and limitations of overlapping consensus, let me conclude with the following remarks. Any party to the overlapping consensus is required to endorse a basic political and democratic ideal of the person as free and equal. Rawls's theory is therefore not addressed to those who do not accept this basic conception of the person, and some 'religious conservatives' may well fall into this category of the excluded. Within the limits posed by the democratic ideal of the person, citizens within an overlapping consensus are free to elaborate more specific and narrowly focused interpretations of freedom and equality. Radical feminist criticism is therefore not discouraged. The obstacles to gender equality may not come from 'religious conservatives' –some of which may be excluded, but from the other sources of political and/or theoretical conflict that remain in a democratic society, even when an

overlapping consensus does exist. A political conception may be able to deal with gender inequality in some way, for instance, by protecting the inalienable rights of women and children, and by setting out measures designed to compensate women who bear the burden of the traditional division of labour.⁷² Political philosophy can contribute towards gender equality, and a social system can be criticised by the principles of justice. However, a political theory cannot effectively deal on its own with a gendered social system. Social theory and human psychology also have a part to play in advancing gender equality.

d) overlapping consensus: practical or normative?

I mentioned at the outset that in laying out the objections to overlapping consensus, I was to follow a particular order of business. My aim was to start with the objection that I considered to be the least damaging to the notion of overlapping consensus and to conclude with the most damaging. Hence I must acknowledge here that this fourth objection can only be partially addressed at present since there shall remain further questions that will be taken up in chapter 4.

To attempt to answer this objection we need to ask whether the role of overlapping consensus is only a practical one (stability), or whether it performs a justificatory function. If overlapping consensus is concerned in some way with justification, then the theory needs to develop beforehand some stabilizing constraints. We know, by Rawls's own admission in his reply to Habermas's objection, that overlapping consensus is concerned with two types of justification: full justification by individual citizens and public justification by political society.⁷³ I leave aside for now the discussion of the particular details regarding these two types of justification.

⁷² IPPR, pp.595-601.

⁷³ PL, pp.386-88.

My aim is to elucidate whether the theory uses the fact of consensus itself as its main basis of validity.

Larry Krasnoff has provided a response on behalf of Rawls that seems fairly persuasive but that is not without problems.⁷⁴ He starts by suggesting that the ideas of stability and consensus play different roles, showing not only that a normative account can be sensitive towards practical problems, but that, in view of the fact of pluralism, it must be sensitive towards practical goals. Stability and consensus 'are deployed in a serial fashion, in a way that reflects a developing normative attitude toward that practical problem.'⁷⁵ His main claim is twofold: a) political justification must fulfil a consensus condition and therefore the object of justification must be available to consensus, and b) the argument for stability steps in to show that this necessary condition can also be a sufficient condition.⁷⁶

With regards to consensus, we know that Rawls's preoccupation lies, not just with any type of disagreement, but with reasonable disagreement. The exercise of rationality in contemporary societies is subject to the burdens of judgement. These are not a product of the limitations of reason but of the fruitful exercise of reason under free conditions. This is why the category of the reasonable is central to political justification. A justified normative theory must meet the criterion of reasonableness, that is, it must be in principle acceptable to all. Although Krasnoff does not mention it, this seems to connect with Rawls's idea of pro-tanto justification. But, as Rawls admits, pro-tanto justification is incomplete since it may not prevail over citizen's non-political views and values. It is at this point that the argument for stability comes in to show that the rational can fully endorse the reasonable. In other words, stability

⁷⁴ Krasnoff, L., 'Consensus, Stability and Normativity in Rawls's Political Liberalism', *The Journal of Philosophy*, 95 (1998), pp.269-292.

⁷⁵ *Ibid.*, p.271.

⁷⁶ *Ibid.*, p.270n.

is concerned with whether citizens as rational beings can accept a theory that has already been pro-tanto justified.⁷⁷ For Krasnoff the crucial point here is that the settling of the question of stability is not a step undertaken by the normative theory itself, but by the various comprehensive doctrines according to their own principles and values. All that the normative theory need do is to define the conditions of endorsement and to explain how general endorsement, that is, an overlapping consensus, can be possible.⁷⁸ Rawls explains this in his discussion of how the political conception and the different comprehensive doctrines may be related. Hence whereas the problem of political justification has been traditionally understood in terms of rational justification, Rawls understands it as reasonable justification which he uses 'as a political proxy for rational justification'.⁷⁹ This requires that comprehensive doctrines show how their particular accounts of rationality support the reasonable as deployed by Rawls. There are three possibilities here: 1)the rational and the reasonable may be identical as in the case of the various comprehensive liberalisms, 2)the rational may be thinner than the reasonable as in the case of the pluralist view, and 3)although the rational may be thicker than the reasonable as in the case of religious doctrines, the reasonable may be required as a condition for the successful exercise of that thicker conception of rationality.⁸⁰ What seems important here, particularly in the context of a reply to Habermas, is that the solution to the stability problem comes as a result of a proper exercise of rationality, and not as the product of practical accommodation. The reasonable is not a practical device used to secure

⁷⁷ In Krasnoff's interpretation, the problem of stability 'arises because (...) the notion of the reasonable is not itself a sufficient condition of political justification under conditions of reasonable pluralism. But we go on to allow this necessary condition to serve as also a sufficient condition so that we may avoid a detailed examination of particular comprehensive doctrines. It is this move, and this move only, that makes the problem of stability central to Rawls's theory of justice'. Ibid, p.284. I intend to examine the notion of stability in the next chapter.

⁷⁸ Ibid, pp.279-83.

⁷⁹ Ibid., p.285.

⁸⁰ Ibid., pp.286-88.

consensus, but a category fully sanctioned by the rational. It is the rational that dictates that political justification must be subjected to the constraints defined by the reasonable.⁸¹ Hence, according to this interpretation, looking at the relationship between the reasonable and the rational in this way enables us to see that Rawls has not sacrificed the normative validity of his theory for the sake of stability.

This reply on behalf of Rawls appears rather attractive. However, I can see an important problem with it because it is not certain that Rawls regards the question of stability as external to his normative theory in the way just described. It is clear that the Rawlsian concern with stability does not arise as the result of a mere practical problem, but to claim that the account of stability is external to the normative theory amounts in practice to suggesting that the theory is complete once pro-tanto justification has been established. This in turn endorses a view of the first stage of the theory as self-sufficient for normative purposes. Beyond this and without the need to assess in any substantial way the degree of stability afforded by an overlapping consensus, the normative force of this idea (that nonetheless Krasnoff wishes to preserve) and of the second stage in general appears to have been lost. There is plenty of textual evidence that suggests that Rawls is prepared to modify his theory if the case for overlapping consensus somehow fails, or at least appears too weak to afford the required stability.⁸² Furthermore, although in the second stage he considers the reasonable to be dependent on the rational in the way described above,⁸³ this seems only to match Rawls's description of full justification –an exercise carried out by individual citizens themselves: 'the citizen accepts a political conception and fills out

⁸¹ Ibid., pp.290-92.

⁸² A clear example is the following: 'What if it turns out that the principles of justice as fairness cannot gain the support of reasonable doctrines, so that the case for stability fails? Justice as fairness as we have stated it is then in difficulty. We should have to see whether acceptable changes in the principles of justice would achieve stability...'. PL. pp.65-6.

⁸³ '...justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen's reason, as explained within its own framework.' Ibid., p.143.

its justification by embedding it in some way into the citizen's comprehensive doctrine as either true or reasonable, depending on what that doctrine allows.'⁸⁴ But we still need to account for the third type of justification which entails the public justification of the theory by political society. This takes place when all reasonable citizens have undertaken their full justification of the political conception. What is important to stress here is that whilst the particular content of the doctrines is certainly irrelevant for public justification, citizens do rely on the fact of consensus itself: 'only when there is a reasonable overlapping consensus can political society's political conception of justice be publicly –though never finally- justified' and there is 'no public justification for political society without a reasonable overlapping consensus...'⁸⁵ Hence, the idea of overlapping consensus does not appear to function as a mere hypothesis,⁸⁶ but as a notion whose successful accomplishment seems to be required for the complete justification of the normative theory. Hence it is both a normative and a practical idea. I have already explained the extent to which it is normative but, in which sense is it practical? We have already rejected the idea that being practical needs to be understood as 'accommodating for the sake of the circumstances'. Rawls clearly rejects this and, to be fair, Habermas's criticism is not intended this way. The practical side of overlapping consensus is closely connected with the idea of political legitimacy. When Rawls discusses the political sociology of an overlapping consensus, he appears to be referring to a kind of stabilizing effect produced by comprehensive doctrines and by the social associations that holders of those doctrines form. Comprehensive doctrines 'play a basic social role in making

⁸⁴ Ibid., p.386.

⁸⁵ Ibid, p.388.

⁸⁶ This is how Krasnoff, following Freeman, sees it. 'Consensus, Stability and Normativity', p.285 *n.* I like to think of overlapping consensus as a premise, rather than as an hypothesis.

public justification possible.’⁸⁷ This social role can be explained in the following terms. Comprehensive doctrines are particular and more or less self-contained bodies of thought that have gained widespread support. A substantial number of citizens in modern societies do belong to associations that are formed on the basis of these comprehensive doctrines and they have a role in developing and propagating them. But most doctrines are survived by the citizens who affirm them. They are therefore contrasted exercises of reason that have stood the test of time. Given pluralism, it is crucial that a political consensus among such doctrines obtains because it is the *only basis of social unity available to us*, and social unity is necessary for stability: ‘in a democratic society marked by reasonable pluralism, showing that stability for the right reasons is at least possible *is part of public justification*.’⁸⁸ Thus showing how the rational affirms the reasonable is part of the justification of the theory (of its public justification) which, in turn, belongs to the realm of normative theorizing. In seeking full justification, we need to check carefully that the political conception of justice does not fatally conflict with, or even reject, the rest of the realm of values that citizens also affirm. Again, this seems a practical aim embedded within a normative approach, and crucially, the successful settlement of the practical problem of stability is required by the liberal principle of legitimacy so that the necessary exercise of coercive political power can be acceptable to all citizens conceived as both reasonable and rational.

We are now in a position to conclude our assessment of the initial question posed by Habermas. This was the question of whether Rawls has given up on normative theory to concentrate on stability issues. Particularly important was his suggestion that Rawls’s definition of the conception of justice as political was

⁸⁷ PL, pp.389-390.

⁸⁸ Ibid, pp.390.

intended as a way of amalgamating two different aspects –acceptance and justification- through the concepts of consensus and stability. In reply to Habermas’s objection, I would accept Krasnoff’s point about the way the two concepts are used in Rawls’s theory, i.e. following a carefully worked out sequence from pro-tanto to public justification. However, I would nonetheless claim that, since the problem of stability and its particular settlement is also part of the normative theory itself, Rawls needs to provide a stronger link between how his conception of justice and comprehensive doctrines are related. In Habermas’s words, he needs to explain ‘the stabilizing effect of an overlapping consensus (...) in cognitive terms.’⁸⁹ We know that, in Habermas’s view, this poses a fundamental problem for Rawls since his decision to define comprehensive doctrines themselves as true prevents him from undertaking just this kind of exercise. My aim here is not to examine all the main differences between Rawls and Habermas, so let me just point to a key final question. We have said that the theory needs to establish more rigorous links between comprehensive doctrines and the political conception. As it stands, it seems that overlapping consensus is a contingent matter. Suppose then that an overlapping consensus on justice as fairness, for whatever reason, cannot be obtained. Rawls’s answer to this is to modify the theory, or to look for another liberal theory of justice, so that it can command widespread support. But, suppose that finding a liberal theory that enjoins general support under the conditions defined by political liberalism proves to be a very difficult task. In this case, we would need to acknowledge that an overlapping consensus cannot be achieved at least in the foreseeable future. But we still need to give the problem of stability an adequate solution since this is required, as we have seen, by the liberal idea of political legitimacy. How this is done in the

⁸⁹ ‘Reconciliation Through the Public Use of Reason’, p.122.

absence of an overlapping consensus is something that cannot be discussed here. But it seems clear that this problem leads us to the question of utopianism.

e) too utopian a consensus?

We have just suggested that an overlapping consensus on Rawls's political conception may not obtain and, if overlapping consensus does not obtain, the problem of stability would not have been satisfactorily addressed in his own terms. Rawls does not really work out the problem of utopianism in a rigorous fashion. All he does is to suggest how we may think of an overlapping consensus evolving from a *modus vivendi*. Instead of an argument, he provides a more or less plausible story of an evolution. In this section I intend to show why an overlapping consensus may be utopian, and to make some suggestions as to how this problem might be overcome.

In seeking an overlapping consensus, it is necessary that people agree not only on Rawls's two principles of justice, but also on his conception of reasonableness and on his idea of the burdens of judgement. He dispenses with the category of truth altogether so that a possible conflict within reason itself may be escaped. The question of truth is something to be considered by the holders of comprehensive doctrines themselves and not by the normative theory. Rawls suspends judgement on the truth or falsity of comprehensive doctrines and instead assumes their reasonableness. This assumption is derived from the idea of the burdens of judgement. Rawls proposes that we view the rational as subjected to specific constraints that lead people to disagree. In acknowledging these limitations people come to accept that any exercise of reason is often influenced by a host of contingent matters, and so they realize that other people's comprehensive doctrines are also valid applications of rationality, even though they may not find them persuasive for themselves. The conclusion drawn from

this is that the features of rationality and its exercise under the free conditions enjoyed in modern democracies prevent us from obtaining any kind of comprehensive ethical agreement.

The next step is then to see whether we can find a satisfactory political agreement. The easiest way to do this would be to discern which type of political consensus people are willing to subscribe to. But the condition here is that the required consensus must be moral, both in its object and its ground. The first requirement is relatively unproblematic for a political theory that aims to be normative since a consensus on something other than moral principles would not satisfy the necessary conditions of normativity. The second requirement derives, as we have seen, from the burdens of judgement. Hence, comprehensive doctrines are not only regarded as viable and more or less rigorous exercises of thought, but they also afford moral credentials to the consensus. With this Rawls seems to have collapsed the distinction between what is reasonable and what is moral. I will come back to this problem at the end of this section. Now I would like to show how the burdens of judgement make an overlapping consensus too utopian.

Let us take the example of the religious view and of how it relates to the political conception. We said above, following Krasnoff, that in this case the rational is thicker than the reasonable but that the reasonable is required as a condition of the successful exercise of that thicker conception of rationality. The type of religious believer that I have in mind, call her B, is one that could accept the two principles of justice together with the idea of the person as free and equal. Moreover, my prototype of religious believer does accept the liberal principle of legitimacy which defines important constraints on the exercise of political power, that is, B submits to a liberal conception of political authority. But the successful accomplishment of an overlapping consensus

demands more than this: it demands that the political conception is embedded into comprehensive doctrines 'all the way down', and this entails the acceptance also of reasonableness and of the burdens of judgement.⁹⁰ In Rawls's discussion of how the political conception may relate to different comprehensive doctrines, he states that 'the acceptance of the political conception is not a compromise between those holding different views, but rests on the *totality of reasons* specified within the comprehensive doctrine affirmed by each citizen.'⁹¹ In the case of religious doctrines that possess an account of free faith, the principle of toleration provide the basis for endorsement of the basic liberties of a constitutional democracy, and therefore, religious believers, on the basis of toleration, can affirm the political conception.⁹² But this link needs further examination. In working out the connections between the religious believer and the political conception, we need to look at whether doctrine holders, rather than doctrines as such, can endorse it. As the name indicates, comprehensive doctrines, and particularly religious ones, are more or less self-contained bodies of thought. Most doctrinal questions are, on the whole, irrelevant for political purposes. At most, religious comprehensive doctrines with an account of free faith provide loose guidance on how the faithful may make sense of their times from their particular religious perspectives. It is then doctrine holders, rather than doctrines as such, who may or may not endorse the political conception.⁹³

⁹⁰ Leif Wenar has raised a similar objection. His critique, however, is motivated by particular aim: to show that these features overburden the theory, are unnecessary for endorsement of a liberal conception and alienate potential support, 'An Internal Critique', pp.41-48. Krasnoff, whose paper is in part a reply to Wenar, suggests that Wenar's interpretation assumes that Rawls's main purpose is to work out an agreement among existing comprehensive doctrines, 'Consensus, Stability and Normativity', p.279n. My aim is to show the internal tensions between different parts of Rawls's theory.

⁹¹ PL, pp.170-71, emphasis added.

⁹² Ibid., p.145 and JFR, p.191.

⁹³ Wenar's discussion, on the other hand, focuses on the doctrines themselves, in particular on whether Catholicism can affirm the conception of justice 'all the way down'. He draws on Catholic doctrinal documents and pastoral letters to illustrate that Catholicism has problems with some elements of Rawls's conception, especially with the burdens of judgement. See 'An Internal Critique', pp.42-57.

Let us say that our religious believer B does in fact affirm a liberal political conception. The support given by B to a liberal political conception comes from: 1) B's acceptance of a liberal understanding of the nature of political authority, i.e. that, given pluralism, political principles must be justifiable to all (ideally), or at least most, citizens and 2) the affirmation of a religious doctrine that regards the person as free and equal in important ways. The way in which B, guided by her particular religious doctrine, understands freedom and equality may differ substantively from the political conception's understanding of freedom and equality. This, however, is not a problem at the outset since Rawls himself constructs his normative theory on the basis of people's general understandings of free and equal so that it can be the focus of a general agreement.

The problem is that although B accepts the fact of pluralism for the purposes of working out the political terms under which she and her fellow citizens have to live, this does not require, in B's view, the acceptance of pluralism as the inevitable consequence of the free exercise of rationality. B, however, accepts that there are different reasonable interpretations of freedom and equality, and different ways of institutionally reflecting this understanding. For B, whereas some types of disagreement are reasonable, others are not so. As a religious believer, B could not affirm that to deny the existence of God is reasonable, although B can accept that to use political power to coerce an atheist into believing is unreasonable. Therefore, in the same way that, in Rawls's account, secular liberals do not have to regard religious comprehensive doctrines as either true or reasonable, religious believers do not have to accept the burdens of judgement 'all the way down'. They may, and can, accept them for reasons of political necessity, but they cannot regard them as philosophically prescriptive for all kinds of matters. Their view of reasonable disagreement is far

narrower than Rawls's; this would not be a problem if the burdens of judgement were not part of the overlapping consensus. But, since the political conception must be endorsed on the totality of reasons specified by the comprehensive doctrines, it appears that religious believers are outside the overlapping consensus. On the other hand, to 'force' people to accept the burdens of judgement entails in practice to 'force' them to affirm a comprehensive liberal doctrine.⁹⁴

There is another possibility unfolding here. Our religious believer B could accept a political interpretation of the burdens of judgement, for this seems a natural conclusion to be drawn from the points just made.⁹⁵ If, after all, Rawls is after a political consensus, why should the consensus entail the acceptance also of the burdens of judgement? But there is a difficulty here. The difficulty derives from the requirement that the consensus must be moral in its ground. If we offer a political reading of the burdens of judgement, the reasonableness of comprehensive doctrines cannot be assumed at the outset. The consensus would lack the moral credentials afforded by comprehensive doctrines. An overlapping consensus is still, however, a moral idea given that its focus, the political conception, is a moral object. A political understanding of the burdens of judgement thus may solve the problem of utopianism, but this may entail the dissolution of the normative into the practical. Without assuming the reasonableness of comprehensive doctrines, an overlapping consensus

⁹⁴ Scheffler asks an interesting question in this respect. The question is whether someone, whom he calls Jane, affirms Rawls's conception of justice and embeds it into a comprehensive doctrine, but fails to regard the conception as political, would be a part of the overlapping consensus. The answer to this appears clear now: if Jane accepts the burdens of judgement and sees other people's comprehensive doctrines as reasonable exercises of thought, she is part of the consensus even if she 'presents them to others as she herself conceives of them, namely, as derived from a certain comprehensive moral outlook.' 'The Appeal of Political Liberalism', p.14. The only condition here is that, when discussing political questions, Jane needs to restrict herself to political values and ideas.

⁹⁵ Krasnoff suggests that without the burdens of judgement the theory is not immune to the charge made by Habermas, even if it could be shown that Catholics, or other religious doctrine holders for that matter, do in fact affirm the two principles of justice, 'Consensus, Stability and Normativity', pp.279-80n. My aim is to show whether a political understanding of the burdens of judgement can make the theory immune to Habermas's criticism, given that it might appear to fend off the problem of utopianism.

lacks the necessary moral ground. This would not satisfy Rawls's conditions for either the full or the public justification of the political conception.⁹⁶

There is another added problem. As we have seen, the moral grounds for both the full and the public justification of the theory are supplied by comprehensive doctrines. There is, however, no reference to the shared moral point of view of citizens as a political body. In the absence of an overlapping consensus, do we then have to fall back on pro-tanto justification? Rawls would not want to follow this route as the liberal principle of legitimacy requires that the political conception be acceptable to all as members of the same political body.

A possible way out of this dilemma is for Rawls to abandon the aim of achieving an overlapping consensus and to concentrate on a different type of consensus, one that offers an account of the political good of the citizenry. This account does not have to be incompatible with people's general accounts of the good since it is restricted to the domain of the political. Perhaps something like Rawls's idea of a constitutional consensus, suitably defined, could serve for our purposes. We would need then to re-work this idea so that the new consensus is also moral in its object and moral in its ground, for nothing less would be satisfactory for the project of political liberalism. Our aim would be to *moralize* the Rawlsian idea of constitutional consensus. Whether or not this re-defined consensus can deal with the problem of utopianism is something that I cannot discuss here.⁹⁷

3.Conclusion.

In this chapter I have surveyed five different objections to the idea of overlapping consensus. I have argued that Rawls can satisfactorily respond to three of these

⁹⁶ Of course we might ask whether this particular account of the moral ground of an overlapping consensus is really necessary. I discuss this in Chapter 5.

⁹⁷ This question is taken up in chapter 5.

objections. The first two critical arguments, namely, that the second stage is redundant and that it rests on suspect empirical grounds, are based on a misunderstanding of the nature of the project of political liberalism, a project whose main aim is to explore the democratic possibilities of justice as fairness. The third objection, that overlapping consensus pre-empts the political, has been dealt with in a different way, by drawing on the implications of Rawls's conception of the person as free and equal. I have also argued that the last two charges can only be partially addressed. There appears to be some difficulties with respect to the aims of stability and justification. If we go along with a political reading of the burdens of judgement so that something like an overlapping consensus might be obtained, it loses its moral force. If, on the other hand, we try to establish more rigorous links between the political conception and comprehensive doctrines, we run the risk of 'utopianism'. Two questions, therefore, remain: one about the relationship between consensus and justification, and the other about the relationship between consensus and stability. Before I can attempt to discuss both questions, I need to examine the notion of stability itself. The next chapter is devoted to precisely this task.

CHAPTER 3

THE PROBLEM OF STABILITY

The primary purpose of this chapter is to examine the idea of stability in both TJ and PL. As we have seen, Rawls's concern with stability first appears in the last part of TJ. The aim of part III of TJ is to show that justice as fairness (the conception of justice already selected on moral and philosophically favoured grounds to well-order the basic structure of society) is able to command the general support of its relevant society¹. A society whose members affirm the philosophically favoured conception of justice can be said to enjoy a, more or less, stable scheme of basic principles of right and institutional practices. Rawls works from the basic idea that providing good reasons for thinking that justice as fairness can in principle engender widespread support is a general requirement for the completeness of his theory of justice². In meeting this requirement, he examines the ground on which the medium- to long-term success of his theory is built upon, in particular 1) the strength of the sense of justice acquired by the members of a society well-ordered by justice as fairness, and 2) the congruence between the principles of right defined by justice as fairness and the general conceptions of the good generally held by citizens. Rawls's conclusion at the end of TJ is that justice as fairness can, all things considered, be regarded as a stable conception of justice because: 1) it promotes a strong sense of justice which is also

¹ See TJ, p.454 and p.391 where Rawls says that we need to see whether "the theory of justice put forward matches our considered judgements and extends them in an acceptable way [and] whether it defines a workable political conception (...). The account of this part is still highly abstract, but I hope to have provided some guidance as to how the principles of justice apply in practice".

² "We have still to complete the theory of justice by seeing how it is rooted in human thought and feeling, and tied in with our aims and aspirations"; TJ, p.391.

consistent with moral learning theory, and 2) it is able to uphold its own principles of right without impinging on peoples' general conceptions of the good.

As we know, this optimism has to a certain extent faded in PL. In his 1993 Introduction he claims that the account of stability given in part III of TJ was unrealistic because it did not acknowledge the fact of pluralism³. Justice as fairness was affirmed on the basis on Kantian morality. Acknowledging the fact of pluralism entails in practice opening up the moral and philosophical foundations for justice, and therefore the very grounds of citizens' endorsement of a conception of justice. According to Rawls, this is a necessary condition for the attainment of stability. Lectures IV to VI of PL deal with this problem and with its proposed solution.

The layout of this chapter is as follows. I begin by providing some general remarks on Rawlsian stability. Then I examine the main features of stability in both TJ and PL. I conclude with some observations on the role that stability plays within Rawls's latter work. A working assumption that will have to be tested is that there may be different conceptions of stability at stake. If this is the case, we need to define them and to evaluate which conception is most adequate for Rawls's political project.

1. Stability: Some general considerations.

Rawls himself admits that the problem of stability has been generally absent from most of the history of moral philosophy⁴. In his view, considerations of stability are crucial if only because they help define the realm of what is practically possible⁵. This is the most basic reason why they should play a more definite role in contemporary political theory. As we have seen, some critics have suggested that in PL the weight

³ PL, p.xix and xlii-xliii.

⁴ TJ, p.455, and PL, p.xixf.

⁵ Rawls quite categorically affirms that "conceptions of justice must be justified by the conditions of our life as we know it or not at all". TJ, p.454.

attached to practical considerations that lies behind the idea of overlapping consensus is now almost as important as the weight attached to normative issues. Most analyses of stability recognize that, given the different aims of TJ and PL, Rawls has given us two different solutions to the same problem. Those analyses, however, do not speculate on the possibility that these two different solutions may be solutions to two different problems. In other words, the question of stability may require a different kind of problematization in each book. One way of looking at this is to examine, what we may call, the particular object or referent of stability in each text. Indeed, underlying those discussions on Rawlsian stability, there appears to be a tacit divergence with respect to the referent of stability. In other words, when Rawls considers stability, is he referring to moral, political or institutional stability? Is stability a property of societies, regimes or theories? And, is stability to be understood in terms of feasibility or legitimacy? Of course, it might well be the case that stability may have different referents, and we need to allow for a certain degree of overlap. In this preliminary section I consider these various referents and meanings of stability. I will not aim to assess them in a rigorous fashion at this stage. This task will be undertaken once the full account of stability has been presented.

A first sense of stability that we might consider is political stability. This is how it has generally been understood by most critics, at least tacitly. In this case, the polity may be regarded as the main object of stability in the following sense: a particular polity may achieve stability if its political institutions, in their day-to-day running, do not depart from the values that ground those very institutions. The particular type of polity that Rawls has in mind is a constitutional democratic state. Rawls's emphasis on the need to reformulate his account of the well-ordered society in view of the fact of pluralism would appear to suggest a concern with this type of

political stability. Affirming justice as fairness on the basis of a single comprehensive doctrine could compromise the political stability of its relevant society because of a likely inequality in the balance of power amongst comprehensive doctrines. Those holding comprehensive doctrines, other than liberalism, will feel underrepresented and this could lead to an erosion of the political project of a society. This could in turn create undesirable tendencies towards instability. So if we accept Rawls's point that justice as fairness in its comprehensive version is not likely to withstand the pressures derived from living in a plural and diverse society, we will need to see if this may have any possible consequences for political stability.

To clarify this, let us consider a society in which everyone is reasonable, and where there is also widespread agreement on the issue of distributive justice. Such a society, one in which a diversity of incompatible yet reasonable views coexist, is unlikely to agree on, for example, the difference principle. It seems reasonable to assume that citizens would hold a variety of views with regards to the redistribution of goods and services. We have then a situation where all views can be said to be reasonable, but one in which, although there is general support towards distributive principles, there is no agreement on one (or one set) of them. So would this create a problem of political stability? The straightforward Rawlsian answer must be affirmative since endorsing the comprehensive view of justice as fairness involves an unquestionable acceptance of the difference principle. Of course, we must grant that the stability of a polity, and particularly of a democratic polity might come about as a result of factors that have little to do with normative theorizing. Most viable polities have different kinds of stabilizing mechanisms regardless of whether they are

democratic and/or just.⁶ Therefore political stability may be, to a certain extent, contingent on social and political circumstances, and only historical analysis can say why some polities are more stable than others. From the point of view of promoting justice as fairness, it might well be an advantage to start with a contingently stable polity because citizens are more likely to cooperate with one another. But Rawls's worry is not whether a particular society is, given its history and context, more or less inherently stable. His worry is that justice as fairness as presented in TJ may itself be a source of political instability under the conditions of pluralism, since 'a continuing shared understanding [of justice] on one comprehensive religious, philosophical, or moral doctrine can be maintained only by the oppressive use of state power.'⁷ Hence, ironically, justice as fairness as a comprehensive view could only aspire to become a stable conception if it was enforced. But then, to what extent could this enforced conception be said to be democratic and just? Hence considerations of political stability do play a part in Rawls's overall account, particularly in PL.

Institutions may be a second referent for stability. Rawls's emphasis on the basic structure of society as the main focus of justice would seem to give some credence to this understanding of stability. A polity can be considered institutionally stable when its citizens agree to adopt a particular set of institutions as a result of a process in which they have participated and on the basis of reasons that they can all accept. When selecting the principles of justice, the parties in the original position know that the principles they are considering are to apply to the main institutions of their society. Hence considerations about how far each set of principles may contribute towards institutional stability may have some weight there. Although in TJ

⁶ This is the view expressed by J. Raz in 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy and Public Affairs*, 19 (1990), pp.3-46, and J. Haldane in 'The Individual, the State and the Common Good', *Social Philosophy and Policy*, 13 (1996), pp.59-79.

⁷ PL, p.37.

some importance is given to institutional stability at the outset,⁸ when discussing stability proper, Rawls seems to conceive this kind of stability as a by-product of moral stability: ‘the stability of a conception of justice does not imply that the institutions and practices of the well-ordered society do not alter (...)...stability means that however institutions are changed, they still remain just or approximately so...’⁹ Hence, in Rawls’s view, moral stability will lead to institutional stability since, in the well-ordered society associated with justice as fairness, citizens’ sense of justice includes a desire to uphold and maintain the institutional arrangements under which they live.¹⁰ In fact, Rawls attaches an institutional constraint when he considers the question of the relative stability of a conception of justice. A conception of justice is said to be more stable than another if 1) its corresponding sense of justice is stronger, and if 2) its corresponding institutional setting is able to contain effectively possible tendencies towards injustice.¹¹ The institutional constraint is thus acting as a barrier against moral instability, but the problem is that Rawls does not go on to assess how and to what extent institutional stability may contribute to moral stability as his discussion is solely focused on moral psychology and on the congruence question. I will return to this matter below when I examine Rawls’s argument more closely.

Institutional stability would appear to be more important in PL. There are two crucial aspects of stability in PL with only the first of these being of interest for us now.¹² Rawls asks us to consider whether people who are raised in a just constitutional setting are likely to develop a sense of justice robust enough to produce general institutional compliance. In other words, moral stability (the sense of justice)

⁸ See TJ, p.458.

⁹ Ibid., pp.458-59.

¹⁰ Ibid., p.454.

¹¹ Ibid.

¹² See section 3 below.

may lead to institutional stability (compliance), but we start with an institutional framework whose role is to promote an effective sense of justice. We do not know whether this just institutional order is stable, but we can presume that it is at least a viable framework of just institutions. Hence whereas in TJ the initial role of institutions was merely to contain injustice (and to contain therefore tendencies to instability), in PL institutions play a more prominent part. The institutional constraint is introduced before any assessment with regards to moral stability and its ability to produce general compliance can take place. Moral stability is still central in Rawls's project, but there is some (if prima facie limited) role reserved for institutions: institutions that are just and stable tend to encourage citizens themselves to be just since 'the institutional form of society affects its members and determines in large part the kind of persons they want to be as well as the kind of persons they are.'¹³

Furthermore, the guiding question of PL –'how is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable religious, philosophical and moral doctrines?'¹⁴ - may suggest a concern with institutional stability. Examining the conditions under which a just society may be stable may involve looking at the strength of its institutions and at how they are bound together. By contrast, in TJ institutional stability is, to a certain extent, taken for granted. What is important here is whether they generate the necessary moral dispositions on the part of citizens.

This leads us to consider moral stability. In TJ stability is primarily regarded as 'a desirable feature of moral conceptions'.¹⁵ Any sound conception of justice must satisfy the 'stability test' if it is to be successful. The 'stability test' is satisfied when such conception can be said to be more in line with the principles of moral

¹³ PL, p.269.

¹⁴ Ibid., p.xxvii.

¹⁵ TJ, p.455.

psychology than its rivals. In chapter 8 Rawls discusses the relative contribution to stability made by different moral theories. One of the main tasks of chapter 8 of TJ is to show that justice as fairness is the most adequate conception from the point of view of stability because of its closer association with the principles of moral psychology, and that therefore it is able to generate its own support. By contrast, a conception of justice that is unable to elicit an effective sense of justice cannot fulfil the 'stability test'. Moral stability thus entails that our preferred conception of justice is psychologically suited to human inclinations.

Let us assume for now that this link holds, that justice as fairness is the more stable conception of justice because it promotes an effective sense of justice. The question that arises is one about the purpose of having moral stability. After all, we can think of instances where a moral conception has been enforced, and therefore 'stabilized'. So if a conception of justice is to be stabilized without being enforced, such a conception must provide an account of how this is to occur. This turns out to be an account of the relationship between a conception of justice and its relevant society. Without any elaboration at this early stage, it appears that TJ and PL provide two different views of this relationship. Whereas in TJ the account of moral psychology has its own status, in PL moral psychology is tied to justice as fairness understood as a political conception of justice. Moral psychology is philosophical rather than psychological¹⁶

To expand on this: in TJ the account of moral psychology is designed to show that the sense of justice associated with justice as fairness is so strong that people raised in a society well ordered by justice as fairness will generally act in accordance with it. An effective sense of justice is necessary to contain the possible

¹⁶ PL, pp.66-68.

unjust forces and practices that may threaten the stability of the preferred moral conception of justice. Stability is important because instability leads to non-compliance and therefore to injustice. In other words, lack of stability on a moral conception may lead to injustice in the medium to long term. Moral psychology is therefore the bridge intended to link principles with actions.

In Rawls's view, an adequate theory of justice needs to explain 'how moral sentiments influence the conduct of public affairs'¹⁷ by showing that moral principles and the norms derived from them are freely endorsed, and not seen as constraints.¹⁸ It would seem therefore that if the account of moral learning does not prove to be sound, justice as fairness could not be regarded as a workable theory. This does not imply that justice as fairness is not philosophically correct, only that it may turn out to be 'utopian', perhaps unrealistically so¹⁹. Thus moral stability acts as a guarantee of the feasibility of justice as fairness.

By contrast, in PL the account of moral learning has been watered down since moral stability depends on whether or not justice as fairness can be the focus of an overlapping consensus of reasonable comprehensive doctrines. This complex question is, in a sense, the focus of this thesis so I will not be pursuing this matter here, but it is worth bearing in mind these remarks made at the outset.

Considerations of feasibility are therefore central to Rawlsian stability: any justified theory of justice needs to satisfy a 'feasibility' test. The account of moral psychology (and the very notion of stability) are 'not intended as justifying reasons for the contract view. (...) At this point we are simply checking whether the conception already adopted is a feasible one and *not so unstable that some other*

¹⁷ TJ, p.493.

¹⁸ Ibid., p.496.

¹⁹ Of course, there is no mention of realistic utopia in TJ. The furthest Rawls goes to acknowledge a possible problem of 'utopianism' is when he says that if the conception selected in the original position fails the stability test, we might need to re-consider that choice. TJ, p.504.

*choice might be better.*²⁰ Hence, by implication, if the conception of justice selected in the original position fails to be stable enough, the parties might have been mistaken in their choice. This claim is very puzzling, but this seems to be Rawls's suggestion. If a conception of justice has already been philosophically justified and, moreover, selected in the original position as the most adequate conception of justice for its relevant society, in what sense, can this (justified and endorsed) conception render itself unfeasible? We can say that this may happen if the sense of justice associated with that conception fails to become nested in people's motives and desires. Again, it is crucial that the account of moral learning holds.

Now, if stability is a practical matter, instability must be the first of Rawls's preoccupations. To see why, consider the following. Suppose that conception of justice A is the best possible conception for society X from the point of view of theoretical reason. Conception A has been elaborated using appropriate philosophical standards of justification and selected by X's parties behind the original position. Suppose that conception B is second best from that same point of view. It may well be the case that in society X, citizens are more able to meet the demands of B, rather than those of A. Although in the ideal world of theoretical reason conception A comes on top, the citizens of X are likely to fail to live up to its demands. Therefore A is, to a certain extent, unstable. Conception B may prove to be a more feasible, and therefore stable, conception of justice for society X.

Let me put this point more simplistically, but perhaps more crudely. Suppose that in the imaginary ranking of conceptions of justice, B does not come second best but, say, tenth best, but it is still the most stable conception of justice from the practical point of view. We could then conclude that the most stable conception may

²⁰ Ibid., emphasis added.

have some deficiencies regarding its standards of justice. In this case, should conception B be adopted instead? (a solution that Rawls clearly does not favour) Or, should we try to change some aspects of A (the philosophically preferred conception) in order to secure its feasibility, and therefore, its stability? How then can Rawls accomplish this?

There is no satisfactory answer to these questions in TJ where Rawls's position is ambivalent. He does not generally claim that justice as fairness is the most stable conception of justice, only that it is *stable enough*. But even this appears somewhat unclear as this passage illustrates: 'There seems to be no doubt then that justice as fairness is a *reasonably stable moral conception*. But a decision in the original position depends on a comparison: other things equal, the preferred conception of justice is *the most stable one*. Ideally, we should compare the contract view with all its rivals in this respect, but as so often I shall only consider the principle of utility.'²¹ Hence, the problem of stability in TJ is, more precisely, the problem of relative stability, although the term 'relative' itself invites misunderstanding. Does it refer to whether justice as fairness is, more or less, that is, relatively stable? Or, does it refer to the extent to which justice as fairness is more stable than its rivals? As we have seen, Rawls admits both interpretations.

There is a final sense in which stability may be interpreted. In PL, stability appears closely related to legitimacy. In a democratic society, where citizens are regarded as free and equal, a conception of justice that aspires to be stable, must be acceptable to each and every person. In a democratic society, an adequate conception of justice must provide 'an account of the legitimacy of political authority as opposed to an account of how those who hold political power can satisfy themselves, and not

²¹ Ibid., p.498.



citizens generally, that they are acting properly.’²² Some commentators have argued that legitimacy in PL comes to replace the argument for congruence (the argument presented in chapter 9) in TJ, and that this creates an added difficulty for Rawls (in PL) with respect to stability²³. The argument for congruence is designed to show that the demands of justice can be seen as congruent with those of the good and that, therefore, people’s conceptions of the good need not pose a threat to the stability of justice. Discussions of legitimacy are generally absent from the framework of TJ, but it is possible to regard this connection as plausible by recalling that on Rawls’s claim ‘acting autonomously is acting from principles that we would consent to as free and equal rational beings (...) They are the principles that we would want everyone (including ourselves) to follow were we to take up together the appropriate general point of view.’²⁴

Besides the claim that autonomy is upheld by following the dictates of the two principles, Rawls also claims that acting autonomously implies that people’s moral sense has been acquired in accordance with principles that would have been generally chosen under fair and contingency-free conditions. Hence a Kantian interpretation of justice as fairness leads us to conclude that in accepting and honouring the two principles, people are simply following their nature as free and equal rational beings.²⁵ The consequence is that the definition of the initial situation of fairness and equal representation yields, almost of necessity, the congruence between the right and the good. To see why we need to go back to chapter 7 where Rawls spells out his conception of the good (the thin theory of the good). Here he crucially

²² PL, pp.143-44.

²³ For instance, Brian Barry in ‘The Search for Stability’, p.890.

²⁴ TJ, p.516.

²⁵ Ibid., p.515.

says that we should expect definitions of the good to be *morally neutral*.²⁶ Indeed in TJ the good is defined in terms of rational life plans. The conditions attached to rational plans are that 1) they are consistent with the principles of rational choice and that they are chosen in full view of the relevant situation and after careful assessment of their possible consequences, and 2) they define rational interests and aims for the persons holding those plans.²⁷ Rawls goes on to say that ‘all the theory of justice assumes is that, in the thin account of the good, (...) such variations as exists in conceptions of rationality do not affect the principles of justice adopted in the original position’, and that ‘the conception of justice adopted is insensitive with respect to conflicting conceptions of rationality.’²⁸

The question that might be asked is that, if justice and goodness are, to a certain extent, independent from each other, why does Rawls needs congruence for stability? It is not the place here to assess Rawls’s argument for congruence, but only to explain the possible connections between congruence, legitimacy and stability now and to suggest that congruence may be a weak argument with regards to stability in TJ.

Thus the crucial difference between congruence and legitimacy is that whereas congruence is insensitive with respect to potentially conflicting rational plans, legitimacy is very definitely sensitive with respect to potentially conflicting reasonable moral views. In other words, legitimacy requires that the right and the good are somehow reconciled, whereas the question of congruence, since it is directed towards peoples’ rational plans (and not to their overall moral views) merely requires that rational plans are brought into line with the requirements of justice. I will come back to this important aspect of Rawls’s argument for stability in the next section. For

²⁶ Ibid., p.404, emphasis added.

²⁷ Ibid., pp.408-09.

²⁸ Ibid., p.447.

now, it suffices to say that, with this difference in mind, it is clear that the consequences for stability are likely to differ. In TJ those whose rational plans clash with the principles of justice may have to be coerced. Hence only rational plans that are compatible with justice may be freely followed: ‘...the principles that best conform to our nature as free and equal rational beings themselves establish our accountability. Otherwise autonomy is likely to lead to a mere collision of self-righteous wills.’²⁹ But, what are the forces that may give rise to a lack of congruence and, as a result, to instability? It seems that, since conceptions of the good are reduced to rational plans that are morally neutral, instability may be defined more in terms of a lack of actual acceptance, than of acceptability. In other words, instability may come about as a result of either a motivation deficit or a conflict of interests rather than as the product of principled disagreement. In PL, the issue is somewhat different. A conception of justice may not be stable if it is unable to elicit the support of free and equal citizens holding reasonable views. Conceptions of the good are not morally neutral and coercion, in this case, may be unjustified since it may question the very basis of the legitimacy of justice.

So far we have surveyed the different meanings of stability in both TJ and PL. Before proceeding any further, let me make the following tentative preliminary conclusions on Rawlsian stability:

1. It seems quite clear that the stability of societies may be achieved independently from any kind of normative account of morality and justice. Only historical analysis can tell us why some societies are more stable than

²⁹ Ibid., p.519.

others. Of course, *de facto* stability may be, all things considered, desirable, but it should not be our concern when examining Rawlsian stability.

2. In TJ moral stability is paramount since stability is primarily a feature of moral conceptions. Stabilizing constraints are needed because acceptability in the original position may not result in acceptance in actual societies. Rawls's conjecture is that the stabilization of a moral conception and its sense of justice is likely to contain possible tendencies towards injustice.
3. In PL, political and institutional stability are particularly important. Let us say for now that all the five meanings surveyed (political, moral and institutional stability, feasibility and legitimacy) have a place there.

The primary purpose of this discussion has been to highlight the complexity and multifaceted nature of Rawlsian stability. In a sense, stability comes in many shapes and forms. I hope to have at least suggested that: 1) an appropriate discussion of stability needs to be related to a particular focus (moral, political and/or institutional), 2) the main focus of stability in TJ may differ from that of PL, and 3) any effective treatment of the issue needs to start by firstly setting the key problem and then relating it to its appropriate object. I now turn to examine more closely Rawls's original version of stability.

2.Stability in TJ.

There is an underlying idea running through the whole of part III of TJ that can be briefly stated as follows: regardless of the intrinsic philosophical value of any particular theory of justice, every theory of justice must include an account of how its principles and values are to relate to its society of reference. Such an account gives us

the missing link between justice, on the one hand, and morality on the other. To be sound, a theory of justice (any theory of justice) must provide a relatively robust psychological outline if it is to adequately bridge the gap between justice and morality. Each conception of justice needs to be tied to a particular psychological account, regardless of its philosophical correctness.³⁰ Hence it is worth reminding ourselves at this point that the grounds for justice as fairness have already been presented, and that the aim of chapters 8 and 9 of TJ is to check that the philosophically preferred theory is also one that is morally adequate, and therefore stable.

In TJ a conception of justice can be said to be stable if it is able to generate its own support, for ‘one conception of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations and if the institutions it allows foster weaker impulses and temptations to act unjustly.’³¹ Moreover, when the institutions defined by a conception of justice are themselves just, those who live under them will tend to acquire a corresponding sense of justice, and thus will contribute towards the stability of those institutions. Rawls presents two arguments in favour of the stability of justice as fairness. The first claim is that justice as fairness is psychologically suited to human inclinations appropriately conceived. The second claim is that justice as fairness is congruent with the good because it is regulative of rational plans. His discussion is presented in two stages, roughly corresponding to chapters 8 and 9 respectively. In the first stage he considers the first argument, that is, the strength of the sense of justice acquired by the members of a society well-ordered by justice as fairness. The question of congruence is taken up in the second stage. In this section I firstly discuss the acquisition of the sense of

³⁰ Ibid., p.477, 504.

³¹ Ibid., p.454.

justice, paying particular attention to Rawls's account of moral development. Then I turn to examine the argument for congruence. I finally conclude with a general assessment of Rawlsian stability in TJ.

2.1. The sense of justice and the account of moral development.

Rawls's aim in chapter 8 is to show that justice as fairness is able to generate its own support because it is more in line (than the traditional alternatives) with the principles of moral psychology.³² The chapter generally deals with two related matters: the content of the sense of justice itself and the process whereby the sense of justice is acquired. Surprisingly, Rawls devotes most of the chapter to the second of these two matters. So we need to elucidate at some point why the discussion of this process merits so much of his energies. Here I only provide a tentative explanation that may need to be revised at the end of the section. My explanation points to the question regarding the status of stability within the overall framework of TJ. As we know, stability is a requirement for the completeness of justice as fairness. Although stability is only explicitly discussed in part III, the requirement for stability, or the idea that stability may be a desirable (and perhaps necessary) feature of a conception of justice may have originated earlier on, more specifically on chapter 3 where Rawls examines the role and features of the original position. When behind the veil of ignorance the parties consider alternative conceptions of justice, they regard stability as one of the requisites for favouring a particular one. However, they have no way of determining how stable each of the candidates potentially is. There is no way of knowing the real degree of stability of each conception a priori. Therefore, although the parties consider stability to be one of the basic features of the best candidate, there is no

³² Ibid., p.456.

assessment to be made of actual stability in the original position situation. The parties reasoning is that if a particular option cannot in principle be regarded as stable, it is to be rejected as the preferred alternative. After a candidate has been selected, it remains to be shown that this chosen conception is able to enjoy the active support over time of actual persons in actual society. That justice as fairness is able to generate this kind of support is for Rawls a very real possibility. This is necessary for the well-orderedness of that society. With these remarks in mind, let us now focus on Rawls's account of the process whereby the members of a society that has selected justice as fairness as the preferred alternative, come to acquire the sense of justice associated with it.

Rawls starts with the premise that it is a common feature of human beings to tend to act as justice requires if we have been exposed to its enduring benefits during the course of our lives. We can say that our desire to act justly is arisen by our experiential knowledge of a just order. To the end of explaining how these desires are firstly nurtured, and then fully developed, Rawls elaborates an account of moral psychology and a related view of moral learning. His main claim is that if he is able to show that the two principles of justice are more closely related to the principles of moral psychology than utilitarian principles, the first part of his argument for stability would have been satisfactorily addressed.

The account that Rawls presents is intended as a generalized abstraction and, purposively, he combines elements from two theories of moral learning, social learning theory and rationalist moral theory. His main concern is to show how the moral sentiments associated with the sense of justice defined by justice as fairness are generally developed. Rawls divides the process whereby a person acquires the relevant moral sentiments into three stages. Each of the three stages in turn

corresponds with one of three psychological laws. Let me briefly outline these three stages³³.

First Stage: The Morality of Authority

In the initial stage of moral development, Rawls wants to look at how moral attitudes are nurtured in children. The domain of the learning process here is the family. The reason for this is twofold. On the one hand, the family itself is part of the basic structure and, on the other hand, it is the first domain in the socialization of children. In this stage, a child's acquisition of moral sentiments depends on those who exercise legitimate authority on the child (usually the parents) themselves displaying these same moral sentiments. There are three conditions attached to this process: firstly, the parents must love the child and manifest this love so that the child grows up with a sense of self-esteem and self-worth (affection); secondly, they must set out clear and relatively simple rules that the child is able to grasp and apply (guidance); and thirdly, they themselves must display behaviour that does not contradict those rules (example). As the child is not yet able to understand the bases of norms and regulations (that may include certain prohibitions), the child's disposition to act in accordance with those norms and regulations is based on them being formulated by those authoritative persons whom s/he loves and admires. Morality is thus based on authority, and authority is, in turn, founded on love and affection.

The first psychological law could be stated as follows: as children are not able to grasp the soundness and validity of norms, a child acquires a collection of moral sentiments when s/he actively acknowledges the parental affection that is manifested in the love that the parents have for their child. The morality of authority

³³ Here I follow the account given in Sections 70-72 of chapter 8, pp.462-79.

is temporary; it is only necessary given the special circumstances of children and their inability to make moral judgements. The final crucial point is that, because of these special circumstances and given the potential dangers of basing a morality simply on authority, its scope must be always dependent on the principles of justice.

Second Stage: The Morality of Association

The morality of association is a step further on the person's development of a moral sense. As the child grows up, morality becomes less a matter of following a set of precepts, and more a question of adopting the point of view of others. The domain of this principle is society at large, ranging from the family to the various associations of civil society, even encompassing what Rawls refers to as the 'national community'.

This stage may also start (although not necessarily) in the family where the child learns the moral standards (rights, duties and virtues) that correspond to the child's own position within it. The content of these rights, duties and virtues are now explained and understood, although at first the child or the young person's conduct may still need to be sanctioned, and approved by those in authority (parents, teachers and leaders of various kinds). The content of the morality of association is rather general and, to a large extent, undefined. Moral standards and ideals depend on the particular association and on the individual's position within it. As the young person engages in a variety of activities that require cooperation, the young person learns to follow the moral standards and ideals associated with those cooperative virtues. Soon he or she may learn that the different positions may require a different point of view.

We can say that the second stage of moral development has been completed when the young person acquires the intellectual powers to see the different viewpoints as consistent with one another and as all belonging to the same cooperative scheme.

Once this is accomplished, the individual is able to understand the features of the variety of perspectives (the different plans and motives, wants and ends and their beliefs and opinions) so that he or she learns to take the place of others. Taking the place of others entails an active engagement with the perspective of those others, and this in turn leads to the individual's adjustment of his/her own behaviour accordingly.

The key aspect here is hence to know how we acquire a capacity to empathise with the feelings of others with whom we associate. The second psychological law states that individuals come to acquire an attachment to others when, on entering the various associations of which they are members, they discover that their associates honour their duties and ideals and that they are bound to others by virtue of belonging to the same cooperative venture. Again, the morality of association bears a direct relationship with the principles of justice. In this case, the relationship is more complex. Some of these associations belong to the basic structure of society and the principles of justice already regulate those associations. The relationship is rather looser in the case of those associations that are not part of the basic structure.

Third Stage: The Morality of Principles

In the final stage of moral development, Rawls focuses on how persons become attached to the principles themselves, so that they no longer act from them moved by a social reason (friendship, fellow feeling, acceptance by society), but because they have internalised those principles and the moral sense that they define. As a result of the morality of association where persons learn to follow the moral demands associated with the different positions held, the sense of justice progressively becomes a crucial part of their moral sense. The morality of association, Rawls claims, 'quite

naturally leads up to a knowledge of the standards of justice'³⁴ as dictated by the third psychological law. Just as the first two laws promote certain attitudes and moral dispositions such as trust, confidence in others and love, the third law states that people's realization of the benefits of having lived under just arrangements results in them acquiring the sense of justice: 'the 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 human cooperation.'³⁵ The consequence of having acquired that sense of justice is twofold. On the one hand, it strengthens those institutions that people generally regard as beneficial, and, on the other hand, it promotes people's desire to create new just institutions when necessary and to reform those that may require change.³⁶ The crucial point here is that particular bonds and attachments do not constitute the basis for this highest stage of morality. What matters here is the citizen body as a whole, a body not bound by these kinds of ties. Therefore their shared sense of justice provides that unifying point of view.

In the morality of principles the conception of right defines moral attitudes and sentiments. These are no longer the product of our contingent situation, but depend on the independent point of view described in the original position. Rawls asks a crucial question here: how can we become attached to the moral principles defined by the conception of right? He provides three reasons to explain how we can do so. Firstly, the content of the sense of justice itself has already been selected by rational persons and, as such, it represents a generally accepted way of settling

³⁴ Ibid., p.473.

³⁵ Ibid., pp.473-4.

³⁶ Ibid., p.474.

possible disputes given the existence of a variety of ends and rational plans. Secondly, the sense of justice, Rawls argues, is compatible with the love of mankind. The same conception of justice applies in both cases, with the only difference being one of focus and demand. The sense of justice is a particular, and more demanding, subset of the love of mankind. The final reason is connected to the Kantian interpretation of the two principles. Given that the principles belong to general conceptions of the good, it follows that people are led to act from them, thus expressing their nature as free and equal rational beings. Rationality therefore leads people to act justly. According to Rawls, the close relationship between the right and the good implies that the sense of justice associated with justice as fairness can be regarded as *natural*.

The three reasons presented in support of the claim that we are likely to become attached to the moral principles defined by the conception of right, also serve the purpose of refuting the doctrine of the purely conscientious act. This doctrine claims that an action in order to be regarded as a moral action, must be primarily motivated by the right itself. Our attachment to what is right and just, and no other reasons external to them, is the source of our moral actions. The only proper moral desires are those that prompt us to act for the right's own sake. A morality that focuses on reasons other than the moral object itself is weak. In Rawls's view, the doctrine of the purely conscientious act borders on arbitrariness and, most importantly, cannot serve as regulative of the basic structure: 'the desire to act justly is not, then, a form of blind obedience to arbitrary principles unrelated to rational aims.'³⁷

Finally, we need to note that the morality of principles also encompasses the other two preceding moralities. In this final stage of moral development all the

³⁷ Ibid., p.476.

previous virtues and ideals are incorporated and organized into a coherent body of general principles. The contents of the moralities of authority and association are thus revised and suitably integrated in the overarching conception.

I have presented this outline for the sake of completeness. I shall now try to extract the main claims made by Rawls here, and analyse them appropriately. Let me first make the following observation. This section of the book has generally received very little attention by critics. Those commentators who have focused on Rawls's view of moral development have bypassed this crucial part; in fact, the account given in chapter 8 has generally been taken at face-value. My aim here is also to show that this section is important within Rawls's account of stability as a whole. Let us now turn to the main claims made by Rawls in his discussion of moral development:

General Claims:

1. The account of moral development is cumulative, latter stages come about on the bases of previous stages.

Claims regarding the Morality of Authority:

1. In the first stage, morality depends on *legitimate* authority.
2. Provided the conditions of affection, guidance and example are met, children develop the required moral sentiments that dispose them to act in accordance with norms and regulations that are suited to their station.

Claim regarding the Morality of Association:

1. When people recognize that different positions imply different points of view, with all points of view belonging to the same cooperative scheme, they learn to empathise with others, and thus to put themselves in the position of others.

Claims regarding the Morality of Principles:

1. The successive adoption of the perspective of others leads to the 'internal' knowledge of the moral point of view.
2. Since just social arrangements have promoted the good, people develop a desire to follow the principles of right because they have benefited from that just social order.

a)The cumulative character of moral development

The first claim suggests that if the sense of justice is to be internalised and become a part of people's overall moral framework (in a morality of principles), it will only happen if the values of the preceding two moralities have been effectively acquired.

The values of the moralities of authority and association form the building blocks, so to speak, of the morality of principles. In other words, persons would not be able to follow the moral precepts and demands defined by the sense of justice unless they have learnt beforehand to follow the values of these other more elementary moralities.

If this is the case, there has to be a kind of internal relationship between the three stages of moral development and their respective domains of values that forms the basis of the moral learning process. I propose to look at this relationship in the following way: a cumulative view of moral development requires that there be 1) a

certain degree of empirical adequacy in each stage, 2) relevance between the contents of the three stages, and 3) consistency between the values and sentiments to be learnt across the overall framework.

Let us start with the question of empirical adequacy. It may seem strange that a normative theory should be concerned with empirical matters but Rawls, although not explicitly, does place some importance upon them. For example, when he claims that it is *natural* for parents to love and care for their children, he is resting on some kind of probabilistic calculation based on a more or less safe assumption. If we take the set of parents in a given society, those who love and care for their children generally outnumber those who do not. Those who do not love and care for their children could provide a source of instability. Since those children would be unable to acquire the required sentiments in the morality of authority, it is unlikely that they come to learn, later on in life, the more demanding values attached to the sense of justice. Of course, it is difficult to measure love and affection, but we may somehow measure care. We can say that children whose basic material and emotional needs are met may successfully learn the morality of authority. I will not discuss here whether this assumption (that it is *natural* for parents to love and care for their children) creates any problems for Rawls's account since this is a matter of content that will be taken up below. I am only interested in explaining why the question of 'numbers', that is, of making his account rests on a kind of assumed (but not tried and tested) kind of quantitative evidence, is not a trivial one.

To this end, we need to recall Rawls's remarks on the concepts of equilibrium and stability which, he says, are used in an intuitive way.³⁸ The concept of equilibrium applies to systems and the kind of system that Rawls has in mind is a

³⁸ Ibid., pp.456-7.

social one. A system is in equilibrium when it has maintained its state over time. In such a system no external force is able to affect its position of equilibrium. Most, if not all, social systems are not in equilibrium as they are continuously affected by external forces of various kinds. However, we may be able to say whether the equilibrium of a particular social system tends to be stable, or whether it tends to be unstable.

The equilibrium of a social system tends towards stability when its internal forces are able to contain external forces that may cause profound changes within that system. An equilibrium is unstable, on the other hand, when centrifugal forces lead to further disruption in the system itself. Social systems are thus stable 'if the departures from their preferred equilibrium positions caused by normal disturbances elicit forces sufficiently strong to restore these equilibria after a decent length of time.'³⁹

Now, what are we to make of Rawls's reference to the 'many' equilibrium states of some systems, and to 'internal' forces within systems? Since the relevant system here is the basic structure of society, we may think of these equilibrium points in terms of the various institutions, groups and associations that generally hold a society together. Since some of these institutions (the family, Churches, Universities, etc.) have endured over time, it is plain common sense to think, as Rawls seems to do, that their bases are firmly rooted in society and that, therefore, would form a secure set of internal forces, all pulling in a centripetal direction. So in his apparent quest for empirical adequacy, Rawls plays with an advantage.

The question of relevance is also important for the cumulative process to obtain. By this I mean that the content of the morality learnt, for instance, in the family, needs to be relevant for the learning processes of the second and third stages.

³⁹ Ibid., p.457.

The content of morality in this first stage is defined by parental authority: children learn the values, norms and precepts handed down by their parents. Rawls does not elaborate on this as his focus is not so much the moral content of what is learnt, but the acquisition of moral sentiments and, in particular, the development of the child's sense of self-worth and self-esteem. In fact, the assumption throughout is that parental norms are generally justifiable.⁴⁰

We do not need to question Rawls's assumption that the family is the first domain in the socialization of children and their first school of morality to see problems here. The assumption that parental norms and precepts are justifiable by virtue of being authoritative norms based on love and affection follows from the view of familial duty required by justice. By familial duty Rawls presumably means the duty of parents to provide a safe and secure environment for their children according to their own possibilities and abilities. This is not a problem as such, provided that there is some kind of empirical adequacy, that is, that the set of parents that abide by this duty outnumber the set of those who do not. The crucial problem relates to the content of the norms (or at least of some of the norms) taught within the family.

The norms taught within the family are of various kinds, and all of them can be said to be derived from a familial duty bounded by love and affection. For example, there are some norms that are practically prescriptive like 'Do not cross the road when the light is red'. Other norms are practically exemplary like 'You use your knife and fork like this'. These are the sort of norms that constitute the day to day pattern of parental education and they are, in a sense, unproblematic. There is, however, another set of norms that are derived from particular moral perspectives and that do impinge on children's moral learning and on the different forms that this

⁴⁰ Ibid., p.463.

process may take. Norms such as 'You have to go to Church every Sunday', or 'If somebody hits you at school, you hit back' have the capacity to shape children's moral development in particular directions. The first of these two norms may be seen as affirmative of a particular moral position, while the second could be considered a reactive norm. I will return to these examples again when I discuss the question of consistency. For now it suffices to say that it is difficult to see the process of moral learning as cumulative if the content of morality in the first stage is not (or, more accurately, does not have to be) relevant to the content of the morality of principles.

A possible answer to this may be that the content of the morality of authority only needs to be related to the content of the morality of association. Thus if we take the parental injunction 'If somebody hits you at school, you hit back', there is a chance that whilst the child may have followed that norm in the primitive and temporary first stage, he/she may have, as it were, grown out of it once the second stage is reached (after all, we can all think of school bullies that have subsequently turned into respectable and law-abiding citizens). The problem is that whereas this may have well been the case, there is nothing in the account that prescribes the content of the first two moralities, and whether they are going to be of use in the process of acquisition of a fairly specific sense of justice.

To this it may be said that the two principles of justice act as a discriminatory norm (Rawls claims this as we have seen). But they apply only in a loose way to the moralities of authority and association. My suggestion here is that Rawls's neglect of any kind of prescriptive content in the first two stages of morality may be a weakness, particularly given the cumulative character of moral development. Here I am not suggesting that the moralities of authority and association lack content, but they lack a specific direction. The virtues of being a good son or a

good daughter, or of being a loyal friend may not have a bearing on the virtues of a good citizen⁴¹.

Let us move now to the question of consistency. This is, I should like to argue, the most important aspect of the three. By consistency I mean that the sentiments learnt in the three stages are consistent or, at least, do not conflict with one another.

Again, Rawls is somehow silent in this respect. Consider this understanding of the morality of authority: ‘...the precepts of the child’s primitive morality of authority are usually expressed in terms referring to external behaviour (...) motives and intentions are largely neglected by children in their appraisal of actions.’⁴² For Rawls, the key to children’s development of the moral sense resides in the norms given by authoritative persons; these norms are simple rules that are relatively easy to follow and that do not require a declaration of motives and intentions on the part of parents. We showed an example of those: ‘You use your knife and fork like this’. Similar examples are: ‘Take your umbrella in case it rains’ or ‘Clean your teeth after breakfast’. These are all norms related to external behaviour and, although at some point children may ask for a motive to follow them (their health, safety and social manners, for instance), these are not motives that typically relate to particular moral sentiments. However, we saw that there are other kinds of precepts that, though presented in terms of external behaviour, may require an explanation of motives, and such explanation may have an effect on the moral sentiments and on their acquisition. In the example ‘If somebody hits you at school, you hit back’, both the precept manifested in that external action, and the reasons given in support of that action do

⁴¹ Clearly (and crudely) put, a terrorist can also be a dedicated father or mother. In a less extreme case, the virtue of honesty, for instance, does not preclude those who exhibit it from, say, supporting an authoritarian political regime.

⁴² *Ibid.*, p.469.

seem to have a bearing on the moral sentiments acquired. If the bases for the morality of authority are to be found on the love and affection that parents have and display for their children, it is all the more pressing that parental norms and precepts are bound by certain limits instead of being assumed as justifiable.

The question of consistency hence applies to the relationship between moral sentiments within and across all three stages. Take the family, for instance. The family is regarded by Rawls as a small association. Once children have acquired the rather primitive morality of authority, the next step is their acquisition of the necessary virtues that makes them good sons or daughters. But, on what basis can we say that a particular young person is a good son or a good daughter? And, more importantly, how does being a good son or daughter impinge on the acquisition of the necessary moral sentiments and virtues? In fact, both school and family are very important places in the development of the morality of association and it is not unreasonable to think that the moral sentiments fomented by each of these institutions may conflict. In our example above ('If somebody hits you at school, you hit back') the parental norm of retaliation seems to encourage a particular set of sentiments (anger, mistrust and self-sufficiency). The corresponding school precept ('Aggressive behaviour will not be tolerated and will be punished') might encourage different sentiments, or at least, it will actively discourage those precepts and sentiments encouraged in the family. I will return to this question of consistency in the next section.

b) The family and moral development

This is the only aspect of Rawls's account of moral development that has attracted some attention by critics (particularly by feminists critics), and in particular by Moller

Okin⁴³. As we have seen, in Rawls's account of how one learns to be a moral person, the family is viewed as a school of morality where children learn to be just. However, this process of moral learning may be trampled at the first stage by the injustice of the gendered family. Rather than opening up the question of justice within the family, he has assumed that families are just without providing reasons for this assumption. My aim here is to elaborate and expand of this main point offered by Moller Okin.

The first thing that we need to ask is whether and to what extent Rawls makes such a crucial assumption. In other words, what evidence do we have to say that Rawls assumes that families are just? And, more importantly, if families are just, on what basis can he claim so? Rawls starting point in the morality of authority is that one of the necessary conditions of human life is the generational transmission of moral attitudes. This seems to be an unquestionable fact of human existence as we know it. From this fact, he goes on to claim, we can assume that the well-ordered society contains *the family in some form*, and that, in families, parents normally exercise legitimate authority over children. We do not really know what kind of families Rawls has in mind, but his reference to 'parents' may lead us to think that he is referring to a heterosexual and cohabiting sort of union. Even so, Rawls admits that we might need to review this institution at some point, presumably in order to fit in with the requirements of justice: '...in a broader inquiry the institution of the family might be questioned, and other arrangements might indeed prove to be preferable.'⁴⁴ In any case, children will always be subject to a morality of authority of some kind since they lack the ability to assess the validity and justification of norms and precepts. The problem here is that, since the basis for the morality of authority is love and affection, it is difficult to see how the required moral sentiments could be

⁴³ 'Reason and Feeling in Thinking about Justice', *Ethics*, 99 (1989), pp.229-49, and *Justice, Gender and The Family* (New York: Basic Books, 1989).

⁴⁴ *Ibid.*, p.463.

developed in other types of institutions. What Rawls should have said instead is that the heterosexual and cohabiting union may need to be revised to give the family a more flexible definition. But Rawls seems ambivalent here. At times he appears to lean towards the prototype of a conservative definition of the family. According to this conception, the main functions of the family are the reproduction of the next generation, the physical protection of dependents and, crucially, their moral education. Furthermore, in this view there is also a connection between the effective moral education of children within the family and their future disposition as citizens. A stable society of law-abiding citizens is the product of moral education within stable families. Of course, there are some points of difference between the Rawlsian and the conservative accounts. The conservative conception views the family in far more rigid terms: the stability of society requires the family to be structured along gender lines. By contrast, in the account of the morality of authority Rawls remains silent with respect to the gendered organization of the family and he concedes, as we have indicated, that changes may have to be made. Hence, in Rawls's account, the particular form that the family might take is irrelevant for the morality of authority. The implication of this is that children will acquire the moral sentiments irrespectively of the type of family structure in which they are raised, with the only requirement being the love and affection of those in authority. This may create problems with regards to the relevance and consistency of sentiments between stages.

Hence it seems that, although the family is considered part of the basic structure of a well-ordered society, Rawls does not require families to be just 'all the way down', so to speak. This condition would have entailed a more prescriptive account in the morality of authority (and an assessment of how the sentiments there are connected to justice). There are nevertheless substantial assumptions in the

morality of authority. The account rests on families being more or less stable (in the sense of not being, what we may call, 'dysfunctional'), and on the content of parental norms being justifiable. At the same time, he neglects the issue of family structures altogether. This, I would like to argue, diminishes the prospects of success of his cumulative account of moral development.

c) Moralities of authority, association and principles: how related?

Thus far we have seen how the standard feminist critique claims that Rawls's account is weak right from the very start since it assumes families to be just. I have refined this critique and argued that although Rawls does not claim that families are just, he makes important assumptions about the degree of stability and autonomy of families, and he ignores family structures altogether. The remarks made earlier with respect to the relevance and consistency of sentiments were primarily intended to show that the process of acquisition of the sense of justice is shaped in important ways by the structures in which these sentiments are learnt. Morality does not operate in a structural vacuum. However, the acquisition of the sense of justice would not only depend on the structural arrangements of the family, as Moller Okin claims, but also on the general arrangements of society's basic structure. In order to make my general point clearer, I would like to illustrate my argument with Alex's story:

Alex's parents were keen sailors. He was born and raised in a boat. Before Alex was born, his parents had decided to spend their lives sailing. In spite of this, they did not consider themselves as having abandoned their society of origin. In fact, they kept a regular but distant contact with it through the radio and newspapers and, more importantly, they had negotiated with the educational authorities a specially designed curriculum for Alex. Life was simple in the boat; the two main tasks were the sailing and the daily care of the boat and of the three sailors. Alex's mother and father had an equal input into these tasks. They would take charge of the sailing in turns, and they would also take up the rests of the tasks, including, of course, their son's education. When Alex reached an adequate age, he too would help with all the jobs. Soon he was given the opportunity to participate fully, with an equal share in all the tasks. Because of this upbringing, Alex had learned about

cooperation, reciprocity and about the need to follow rules and procedures. When he reached 18, he decided to study with a distance learning university, and he was eventually awarded both a First degree and a Masters degree.

Soon after Alex had completed his studies, Alex's father fell ill. In one of their visits to port, he was diagnosed with a chronic illness, nothing life-threatening in principle, but something that required constant hospital monitoring. Alex and his parents had to adjust to living in the city again. Alex successfully applied for a middle managerial position with an insurance firm. He was cautious but happy to start an almost new life.

Adjusting to this new life was certainly a difficult business. Certain things in his company did not cease to amaze him. The first thing he noticed was that about 85% of the people with a position similar to his were men. Moreover, there were hardly any women in positions above his. That was strange as he would have expected the numbers of men and women to be roughly similar. The second thing that he noticed was that although the company had rules and procedures in place, the top managers were nearly always right in their decisions, regardless of whether any rules had been contravened. In such circumstances, Alex found it very hard to cooperate and the situation in the firm made him question the values of reciprocity and regard for others that he had learnt with his parents in the boat.

In simple terms, this story exemplifies the shortcomings of Rawls's account of moral development. This account does not only assume, as we have seen, that families are just, but also that society at large, and particularly its background institutions (Churches, Universities and the associations of civil society) are also just, or at least approximately so. Hence, injustices may not only arise in the family; in our example, there was no continuity between the values of equality and reciprocity that were learnt in the family, and the values of a particular association. If the morality of principles needs to be nurtured at the previous stages, there must be some consistency of fit across the three stages. Our concern should not only be the justness of the family, but also the justness of society's background institutions, and particularly of those institutions, like companies, that have a primary role in the allocation of economic positions. In other words, we should look at the implications for justice of public, as well as private, patriarchy⁴⁵.

⁴⁵ Walby, S., *Theorizing Patriarchy* (Oxford: Blackwell, 1991)

To conclude this critical comment on Rawls's theory of moral development, let me suggest the following. Rawls's outline of the process of acquisition of the sense of justice seems primarily aimed at providing a link between the requirements of justice and human motivation, a link that, in his view, is missing in a utilitarian conception of justice. However, what strikes me is that the same account of moral learning could have been provided in support of other theories of justice. I am not saying here that it could have supported *any* theory of justice. I am only claiming that the account is sufficiently loose, and not prescriptive enough, in its connections with the two principles of justice as fairness. Other theories of justice might have deployed a similar view of their stability. After all, Rawls himself acknowledges that most theories of justice view human nature, and the relation between principles and experience in a way similar to his⁴⁶.

2.2.The Idea of the well-ordered society.

In this section, I wish to make some pertinent remarks on the idea of the well-ordered society, and on its connection with stability. We can express this connection as follows: a well-ordered society is a society whose basic structure is regulated by a conception of justice where such conception is: 1)the philosophically preferred conception of justice for that society and 2)a conception that is sufficiently stable⁴⁷. A society can be said to be well-ordered by a conception of justice if: 1)that conception is publicly known and accepted as the preferred conception of justice and 2)the society in question is effectively regulated by it. If we now look at the two conditions as they apply to the well-ordered society of justice as fairness, we arrive at the following conclusions. Condition 1 derives from the contractarian point of view. The

⁴⁶ Ibid., p.456.

⁴⁷ Ibid., pp.453-54.

publicity condition is built into the very definition of the original position⁴⁸. The parties in the original position know that they are to select principles of justice that are to be organized into a public conception of justice. They know the basic general features of their society, and they also know that part of their task is to assess the various conceptions of justice by considering how far they can go in fulfilling the publicity requirement. When behind the veil of ignorance the parties evaluate alternative conceptions of justice, their task is to select one option as follows. Firstly, given that the parties are not only aware, but also take on board, information concerning the relevant features of their society, they consider general notions of justice and common views about the proper relations between persons. Secondly, they assess different conceptions of justice in the light of their knowledge of both these general features and the widespread ideas about justice current in their society. And thirdly, they select the conception of justice that reflects these general ideas most accurately. The publicity condition is thus met at two levels. At the first level, the parties study, or so we can assume, their society's public understanding of justice. Here ideas might range from the more intuitive ones, to the vaguely articulated and then to the more sophisticated ones. At the second level, the parties are confronted with a list of alternatives and they select the one that reflects most accurately the public understanding of justice of their society in its different manifestations. The point to be stressed here is that the publicity condition yields the general acceptability of the conception of justice. In the original position the preferred conception, justice as fairness, is selected on the grounds of its general acceptability.

Condition 2 is less straightforward, for how are we to account for the effectiveness of a conception of justice? It would seem that a society effectively

⁴⁸ Ibid., pp.17-22.

regulated by its philosophically favoured conception of justice would be one whose institutions reflect and uphold the normative underpinnings of such a conception. In Rawlsian terms, it would be one whose basic structure is organized and governed by the two principles. However, Rawls adds a second requirement, that of its members' long-term endorsement of the selected conception. We could say that a society can be said to be simply regulated by justice as fairness when a) the institutions of its basic structure comply with its two principles, and b) when its citizens generally endorse justice as fairness. When this is measured over time, a society can be regarded as effectively regulated by justice as fairness, and so well-ordered. Condition 2 can be re-phrased as the stability condition.

Now there is a further sense in which stability is a requirement of the well-ordered society. If we construe stability as acceptability measured over time, there may be a problem derived from the fact that initial general acceptability is obtained through the contractarian method employed. For, how can Rawls ensure that justice as fairness, the conception selected in the original position, will continue to be the preferred option? Hence, the well-orderedness of a society cannot be guaranteed unless it can be shown that support for justice as fairness can maintain itself over time. Hence, for a society to be well-ordered, its conception of justice must be able to generate its own support. We have already seen how Rawls thinks that this can be accomplished.

Now, at the end of chapter 8, Rawls also offers an explanation of why, in his view, justice as fairness is able to provide relative stability.⁴⁹ His claim is that the sense of justice defined by justice as fairness is stronger than that of other conceptions. The principles of justice reflect an idea of reciprocity where the pursuit

⁴⁹ Ibid., pp.499-504.

of a person's good is secured by the restrictions of equal liberty on the conduct of other persons and institutions. The different priority rules, the Kantian underpinnings of the difference principle and its links with an idea of fraternity strengthen the reciprocity principle. This in turn increases peoples' sense of self-worth and self-esteem which results in 'a closer affiliation with persons and institutions by way of an answer in kind.'⁵⁰ The sentiments that the two principles encourage and the attachments that they produce are stronger than those that are associated with the utility principle.

The strength of the sense of justice is also influenced by the clarity and transparency of the moral conception, particularly when compared with teleological doctrines such as utilitarianism. The contract view offers a kind of simplicity where general claims can be adjudicated by the two principles; it is relatively easy to tell when the equal liberties are met, far easier, in any case, than to make judgements on the kinds of steps that may increase social welfare. The principles of justice also set clear criteria for the types of conduct expected from those who are subjected to them.

Finally, the sense of justice that justice as fairness encourages is based on the attractiveness of its ideals. The main aim of this conception is to provide a way of responding to those contingencies that are morally arbitrary. The removal by political institutions of any barriers to inequality is also required. The assumption here is that 'one of a person's natural wants is that there should be harmony between his feelings and those of his fellow citizens.'⁵¹ But, importantly, this is not effected by a compromise between altruism and egoism, as Mill thought. In the contract doctrine, the same theoretical resources used in support of the two principles are also used in

⁵⁰ Ibid., p.499.

⁵¹ Ibid., p.502.

support of the relative stability of the moral conception. The account of relative stability is 'internal' to justice as fairness itself.

2.3.The Congruence Argument.

The preceding remarks were required for my assessment of the congruence argument. As we said above, the argument for congruence is designed to show that the demands of justice can be seen as congruent with those of the good and that, therefore, people's conceptions of the good need not pose a threat to the stability of justice. This is the second, and final, argument for stability and it is given in chapter 9 of TJ. Rawls's intention here is to strengthen and complete his account of stability for, in his view, 'it remains to be shown that given the circumstances of a well-ordered society, a person's rational plan of life supports and affirms his sense of justice.'⁵² His working premise is that in a well-ordered society the sense of justice is part of a person's good. When this happens, the possible forces of instability are at least contained, when not eliminated. Hence the final chapter of TJ is designed to show how the well-ordered society associated with justice as fairness passes the ultimate test of stability. In such a society, a sense of justice is in effect part of people's good. A society whose basic structure is governed by justice as fairness, and where support for that conception has maintained itself, is a society that is perceived as a good in itself. Its just institutions can be seen to work towards the good of its members. The well-ordered society reflects the idea of a social union when it achieves an idea of community, for 'human beings have in fact shared final ends and they value their common institutions and activities as good in themselves. We need one another as partners in ways of life that

⁵²Ibid., p.513.

are engaged in for their own sake, and the successes and enjoyments of others are necessary for and complimentary to our own good.’⁵³

Two questions arise from looking at the issue of congruence. The first one relates to the internal argument for stability within TJ, whilst the second is concerned with the relationship between stability in TJ and PL. With regards to the first question, we said in 2.2 that, for a society to be well-ordered in Rawls’s own terms, it has to satisfy two conditions: acceptability and stability (understood as acceptability over time). We saw how acceptability arises from the publicity requirement within the original position. The requirement of stability also obtained, based on the soundness of the account of moral psychology. When the two requirements are met, we may say that we have a society well-ordered by a stable conception of justice. For Rawls, the relative strength of the content of the sense of justice (provided by the moral conception of justice as fairness itself) coupled with an effective view of moral learning already secure a well-ordered society.⁵⁴ From this it follows that, given that we already have a well-ordered society, it is difficult to see how the argument from congruence presents a further argument for stability. If this were the case, an increase in stability, as a result of congruence, would produce a parallel increase in the degree of well-orderedness of society. But stability is already a feature of the well-ordered society. The most that the argument for congruence can do is to show that, in a well-ordered society, people’s sense of justice may become a part of citizens’ life plans (hence the idea of congruence between justice as fairness and goodness as rationality). Therefore, chapter 9 merely describes how a society well-ordered by justice as fairness could look like in the long run.

⁵³ Ibid., p.523.

⁵⁴ See, for instance, *ibid.*, p.456: the task of chapter 8 is to show that justice as fairness ‘is likely to have greater stability than the traditional alternatives, since it is more in line with the principles of moral psychology. To this end, I shall describe briefly how human beings in a well-ordered society might acquire a sense of justice and the other moral sentiments.’

The second question is perhaps more crucial because it relates to how stability has been problematized in PL. In his 1993 Introduction, Rawls insisted that the main departures from TJ originated as a response to an internal problem of justice as fairness and an important related inconsistency within his theory as a whole. The problem, as we know, derives from his conception of stability given in part III, and the inconsistency comes from the relation between his account of stability and his idea of the well-ordered society. The idea of the well-ordered society as presented in TJ has made itself unrealistic because it is unable to secure the stability of its own conception under the conditions of modern pluralism.⁵⁵ That first introduction to PL does not contain an explanation of which part of the argument for stability was problematic. We need to wait until his 1996 Introduction for clarification on this. Here Rawls admits that his account required the implicit affirmation by citizens of justice as fairness understood as a comprehensive view.⁵⁶ In the well-ordered society of justice as fairness, citizens endorse at least part of Kant's comprehensive liberalism.

Rawls does not specifically point to where exactly his account of stability faltered, but both Barry and Freeman have suggested that the problems derive from chapter 9 and, in particular, from the argument for congruence. This argument, Barry claims, was designed to close a gap of motivation between citizens' acceptance of principles and their desire to honour them in their actions.⁵⁷ To this end, he might have provided a somehow biased Kantian view. However, by contrast with Rawls's own perception, Barry thinks that this does not turn justice as fairness itself into a comprehensive conception of justice, only the account of its stability. More importantly for our purposes here is his view that overlapping consensus (the account

⁵⁵ PL, p.xix.

⁵⁶ Ibid., p.xlii.

⁵⁷ 'The Search for Stability', pp.883-87.

of stability of PL) originates as a reformulation of chapter 9 of TJ. Similarly, Freeman argues that it is chapter 9 and the argument for stability presented there that pushes justice as fairness towards its comprehensiveness. Chapter 9 has a deeper Kantian basis than the rest of the book since the congruence argument seeks to show: 1)that justice is an intrinsic good, and 2)that justice is the supreme good.⁵⁸ This, Freeman suggests, is what leads Rawls to include comprehensive doctrines within the framework of PL since chapter 9 was biased in favour of public reasons and exhibited a kind of cultural intolerance against non-public reasons. Overlapping consensus, in a sense, tries to restore the balance in favour of non-public reasons.⁵⁹

We need not question here Rawls's own diagnosis of the comprehensiveness of TJ, however, we can cast some doubts about Barry and Freeman's claims with regards to the relationship between overlapping consensus and chapter 9 of TJ. The problems arise from the different understandings of the good deployed in TJ and PL. As we mention in section 1, in TJ references to the good need to be understood in terms of rational plans. The question of congruence can be defined as the consistency between the requirements of justice and the dictates of people's rational life plans. Crucially, Rawls thinks that definitions of the good may be generally regarded as morally neutral.⁶⁰ And, we can see how this differs from the idea of the rational good implicit in an overlapping consensus where substantially religious, philosophical and moral doctrines play a prominent part in defining citizens' rational goods.

⁵⁸ 'Political Liberalism and the Possibility of a Just Democratic Constitution', *Chicago-Kent Law Review*, 69 (1994), pp.619-68, p.628.

⁵⁹ *Ibid.*, pp.632-38.

⁶⁰ TJ, p.404.

2.4. Stability in TJ: Conclusions.

Let me conclude this review of stability in TJ by making the following general observations:

1. This overview of stability has tried to show how TJ's concern with stability arises as a result of using the original position as a device for selecting the principles of justice. The argument for the relative stability of justice as fairness based on reciprocity is not completely separate from the argument for the two principles and it is, to a certain extent, a corollary of Rawls's contractarian methodology. Hence, we can say that a satisfactory solution to stability in TJ is required as a counter-balance to a possible motivation deficit on the part of citizens who, once the veil of ignorance is lifted, may not honour the commitments derived from the initial acceptability obtained in the original position⁶¹.

2. TJ contains two arguments for stability: the account of moral learning and the argument from congruence and these were offered in chapters 8 and 9 respectively. The argument presented in chapter 8, I have argued, is not as strong as it might appear. If part of the problem of stability is that of showing how that initial acceptability may be sustained over time by supplying appropriate motivational stimuli, the account of moral learning seems to falter on two counts. Firstly, even assuming that a close connection may be established between living in just society and citizens' acquisition of a just ethos, Rawls's account is rather loose, and not prescriptive enough. Ultimately, the same view could have been offered in support of other conceptions of justice that contain a similar account of the sources of human motivation. Secondly, the process of moral learning does appear in a kind of social, institutional and, generally, structural vacuum. This is important because, as Rawls

⁶¹ See McClellan, E. F., 'Justice and the Problem of Stability', *Philosophy and Public Affairs*, 18 (1989), pp.3-30, p.4.

seems to acknowledge, structural constraints, particularly of an institutional kind, affect in important ways the scope of moral agency. With respect to the second argument for stability –the congruence argument-, I have tried to show how this, rather than an argument for stability, depicted the effects of stability in a well-ordered society.

3.As I suggested at the outset, TJ is primarily concerned with moral stability since stability is a feature of moral conceptions. Moral stability is seeing as contributing to institutional stability for, ‘stability means that however institutions are changed, they still remain just or approximately so, these adjustments being called for by new social circumstances. The inevitable deviations from justice are effectively corrected or held within tolerable bounds by forces within that system. Among these forces I assume that the sense of justice shared by the members of the community has a fundamental role. To some degree, then, moral sentiments are necessary to insure that the basic structure is stable with respect to justice.’⁶² This shows, rather clearly, that in TJ the attainment of moral stability is a necessary condition for the stability of justice. I now turn my attention to PL. I provide a much briefer review of stability here because, since the framework of PL is the subject-matter of my project, some of the issues surrounding stability will be discussed later in this thesis. My aim is only to examine why Rawls might be concerned with stability in PL given its relaxation of the contractarian methodology.

3.Stability in PL.

We have seen why Rawls might have been concerned with stability in TJ. We also know, from reading his two introductions to PL, that he remained worried about

⁶²Ibid., p.458.

stability. My working hypothesis here –an hypothesis that will need to be tested- is that whilst we should continue to be concerned with stability in PL, the underlying reasons for these concerns differ in important ways from those of TJ. My suggestion here is that stability in PL is an inherited problem (from TJ), but that it may be a different problem (from the one in TJ).

In the two introductions to PL, Rawls offers a diagnosis of the problems of TJ. In his view, the main sources of tension within his theory derive not just from his account of stability, but also from his idea of the well-ordered society. This idea needs to be re-formulated, Rawls suggests, to take stock of the fact of reasonable pluralism and of his shift to the ‘political’. In his 1996 Introduction, Rawls also regards, besides stability and the well-ordered society, the account of moral psychology and moral learning given as problematic on the grounds that it is too Kantian. Generally, there are two problems with the version of stability given in TJ. The first problem is that this version is too utopian, and not realistically utopian since ‘it is inconsistent with realizing its own principles under the best of foreseeable conditions’,⁶³ where these conditions refer to the democratic fact of pluralism. The second problem is that in the well-ordered society associated with justice as fairness, citizens endorse the two principles on the basis of Kant’s comprehensive liberalism. This ignores the fact that constitutional democracies –the type of political order to which the two principles apply- contain a diversity of comprehensive doctrines and, furthermore, this is not an accident or a historically limited condition, but the result of the exercise of reason under free institutions. Hence, Rawls’s dissatisfaction with his previous problematization of stability arise not only from his concern with the (utopian) character of the well-ordered society but also, rather crucially, from reflecting on the

⁶³ PL, p.xix.

key features of democratic societies. There seems to be an unwitting realization that, even though in TJ the principles of justice were designed to apply to the institutions of the basic structure of a constitutional democracy, the democratic character of such a system was taken for granted, rather than appropriately established and its connections with justice as fairness worked out in detail. We may even say that justice as fairness as presented in TJ suffered from a kind of democratic deficit. The main problem being that it did not acknowledge the extent of pluralism.⁶⁴

But here some questions loom. For instance, why does the shift to the political require a re-definition of the well-ordered society? What are the features of his new conception of the well-ordered society? And, how exactly do the changes in the theory (marked by the shift to the political) call into question the account of stability of part III of TJ? These are awkward questions, not least because Rawls himself appears somewhat elusive on these matters. We need to find an explanation of why these ideas as conceived in TJ do not survive the move to PL. It is interesting to note that, even though Rawls acknowledges the centrality of the question of stability within the project of political liberalism, his explicit discussion of stability only covers one section (section 2) of one lecture (Lecture IV)⁶⁵. Rawls offers there his most complete problematization of stability in PL⁶⁶.

As we know, the argument for stability in PL is divided into two parts. In the first part, justice as fairness is elaborated as a free-standing political conception. The question of stability only arises in the second part. Here Rawls takes up the question of whether justice as fairness can be considered sufficiently stable for, 'unless it is so, it is not a satisfactory political conception of justice and it must be in some way

⁶⁴ As we have seen, pluralism only covered the case of rational life plans.

⁶⁵ *Ibid.*, pp.140-44.

⁶⁶ Although this is expanded in JFR, pp.180-202.

revised.⁶⁷ On the basis of what was said above, justice as fairness would need to be revised if: 1)it proved to be unrealistically utopian, or 2)it continued to sidestep the democratic features of the societies to which it applies.

In PL stability is concerned with two separate questions. The first of these is similar to the question of TJ –whether people brought up in a just society can develop the required sense of justice to comply with the society’s just institutions. The second question is specific to PL; it asks whether, in the face of pluralism, the political conception can be the focus of an overlapping consensus. The answer to the first question rests on an account of moral psychology –an account that is philosophical, not psychological.⁶⁸ It is based on the capacities and features of citizenship, and on the connection between people’s two moral powers and their desire to exercise their ideal of citizenship. Citizens are assumed to have a reasonable moral psychology, and this involves an active acceptance of the fair terms of cooperation, acknowledging the burdens of judgement, endorsing reasonable conceptions of justice, and showing their commitment to being a member of the citizenry.⁶⁹ Answering the second question of stability requires setting out the circumstances under which the attainment of an overlapping consensus may be possible. This second question of stability provides the basis for the main arguments in this thesis, and hence, we need not elaborate on this here. With regards to the first question and its corresponding answer, all we need to say is that Rawls’s view of moral psychology represents an improvement from the view of TJ because it is connected to a specific political ideal such as democratic citizenship. My next chapter will deal with some related problems arising from Rawls’s conception of the reasonable person.

⁶⁷ PL, p.141.

⁶⁸ Ibid., pp.81-88.

⁶⁹ Ibid., p.86.

I turn now to a crucial matter. Stability, Rawls claims, is different from simple feasibility since stability cannot be achieved at all costs. If a liberal political conception of justice, such as justice as fairness, were to become stabilized as a result of oppression, enforcement, or even persuasion, it would cease to be a liberal political conception. A liberal conception must relate to stability in a qualitatively different manner. A liberal conception must be formulated in such a way that it can elicit support from each and every citizen as reasonable and rational: 'justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen's reason, as explained within its own framework.'⁷⁰ This kind of stability also defines the conditions under which political power may be legitimately exercised, for only so is it an account of the legitimacy of political authority as opposed to an account of how power can be exercised by those in authority. Here Rawls ties the notion of the legitimacy of political authority to the attainment of stability through overlapping consensus. We saw in the previous chapter the problems surrounding overlapping consensus. Hence, if overlapping consensus does not in the end obtain (and, as we have seen, Rawls himself clearly admits this possibility), what are the implications of this for legitimacy? If overlapping consensus cannot be guaranteed, and it seems that it cannot, how does political liberalism account for the legitimacy of political authority?

To be certain, legitimacy requires a different treatment from stability, for the question of legitimacy –what are the bases of citizens' allegiance to the political authority that they are to submit to?-, differs from the problem of political liberalism: how is it possible that there may exist over time a stable and just society of free and

⁷⁰ Ibid., p.143.

equal citizens profoundly divided by reasonable and incompatible doctrinal conflict?⁷¹ Conflating both questions appears deeply problematic because legitimacy is made to rest on a highly demanding condition. If legitimacy requires overlapping consensus, then it also requires a concordance of fit between the public and private moralities of each and every citizen (or at the very least of most citizens).

However, Rawls's statement of the liberal principle of legitimacy - 'our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may be reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason'⁷²- does not mention overlapping consensus, but only consensus on the essentials of a constitution. This may have its difficulties, but it is a substantially weaker notion of legitimacy. I will pursue this matter in chapter 6. For now it suffices to stress that, in PL, Rawls is concerned with two separate questions; the first question is whether, and to what extent, justice as fairness is able to stabilize a constitutional democracy 'for the right reasons'. The second question focuses on how liberal political philosophy must understand the appropriate basis of citizens' allegiance to political authority. The answer to the first question is provided by the idea of overlapping consensus, whereas the answer to the second question is, as we have seen, ambiguous.

There is one final point to consider. We mentioned above how Rawls's own diagnosis of the problems of TJ focuses on the idea of the well-ordered society. This idea is too utopian and needs to be revised. However, as Barry has pointed out, the definition of the well-ordered society given by Rawls in PL has not changed

⁷¹ Ibid., p.xx.

⁷² Ibid., p.136.

substantially from that of TJ.⁷³ This leads Barry to claim that justice as fairness was already political in its first presentation. However, whilst the concept of the well-ordered society may not have changed, its conception may have done so, for now a well-ordered society must be a democratic one: ‘any conception of justice that cannot well-order a constitutional democracy is inadequate as a democratic conception.’⁷⁴ Hence, even though the requirements for well-orderedness may not have changed, a society cannot be well-ordered unless it is democratic, for Rawls is focusing exclusively on the particular case of a constitutional democracy. Therefore, the conception of the well-ordered society here is qualitatively different from the one presented in TJ.

4. Conclusions

This chapter has provided an analytical account of stability in both TJ and PL. I started with the underlying assumption that, even though Rawls has used the same concept of stability, this concept may have corresponded with different conceptions. This was confirmed by examining critically the different questions that led Rawls to problematize stability in each book. We saw that in TJ stability originated, in an important respect, as a response to a kind of motivation deficit derived from its contractarian framework. Also, the main concern in TJ was with moral stability and its relative contribution to political stability. In TJ the argument for stability is presented only after the justification of the moral conception of justice to which it applies. Then we saw how the problem of TJ was not so much the comprehensiveness of its account of stability, but the insufficient attention given to the democratic fact of pluralism. Acknowledging the fact of pluralism requires a different kind of treatment

⁷³ ‘The Search for Stability’, pp.879-80.

⁷⁴ PL., p.35.

of the two key questions stability and justification. This effects a division of the framework of political liberalism into two parts, with the first part focusing on justification, and the second part on stability. This division, however, is not clear-cut and questions of justification also arise at the second stage. But PL was also concerned with a third issue, that of legitimacy. The question of legitimacy differs in important ways from stability. Legitimacy focuses on the grounds for citizens' general allegiance to a political authority. As we have seen, Rawls conflates the two issues of stability and legitimacy, and this has important consequences for the standing of his project as a whole. Given the difficulties faced by overlapping consensus, both stability and legitimacy appear to prevent justice as fairness from being a realistic utopia.

Even though the account of stability in TJ exhibited some important weaknesses, such weaknesses are of a different nature than the difficulties faced by stability in PL. TJ presents an account of the moral stability of justice as fairness from its own perspective. Justice as fairness is, we may say, self-referential.⁷⁵ Failure to achieve stability may not necessarily entail the revision of the theory, but only of the account of stability itself. Rawls's main aim is to show how a moral conception may be stabilized in order to keep tendencies towards political stability in check. The problems of PL are of a different nature. The purpose here is to show how to achieve political stability on moral grounds ('stability for the right reasons'). Stability is not a feature of moral conceptions since the focus of stability, justice as fairness, is now regarded as a political conception. However, he still makes important concessions to moral stability. If moral stability is attained when the majority of active citizens in a society have internalised the political values that are embedded in their society's basic

⁷⁵ This does not necessarily mean that it is comprehensive.

institutions and have incorporated them into their respective domains of values, both stability and legitimacy seem to require this kind of moral stability. As a way out of this dilemma, I propose to focus on institutional and political stability. A polity can be regarded as institutionally stable when its citizens agree to adopt a particular set of institutions as a result of a process in which they have participated and on the basis of reasons that they can all accept. This kind of stability does not require a concordance of fit between citizens' private and public moralities, but only 'institutional' consensus, albeit 'for the right reasons'. Chapter 5 will be given over to an assessment of institutional stability. Chapter 6 will be concerned with political stability in relation to both stability proper and legitimacy. Political stability is attained when a society's political institutions, in their day-to-day running, do not depart from the values that ground those very institutions. I will show how a society may achieve political stability. In my view, a political project such as Rawls's should be exclusively concerned with these two types of stability, for the settlement of moral stability is a step within the scope of a moral theory itself, comprehensive or otherwise.

CHAPTER 4

PLURALISM AND POLITICAL LIBERALISM

After having examined the question of stability and pointed to some crucial difficulties that political liberalism faces, I turn now to the second of our problematic notions, that of reasonable pluralism. One of the challenges that we face when dealing with this notion is that nowhere in PL can we find something like a rigorous working out of this central aspect of Rawlsian thinking. All we find are a series of remarks distributed across different parts of the book.¹ My first priority is therefore to offer a reconstruction of Rawls's account of the fact of pluralism. Once this task is accomplished, I proceed to consider some possible weaknesses that it may face. My general aim in looking at these possible shortcomings is to see whether, and to what extent, they affect the realization of Rawls's project in any substantive way.

The first reference to the fact of pluralism can be found in IOC. The fact of pluralism is connected with the existence in modern democratic society of a diversity of doctrines – a diversity of general and comprehensive moral views that in turn nest a variety of conceptions of meaning, value and purpose of human life and activity.² Such general and comprehensive conceptions are both conflicting and incommensurable,³ and therefore their respective aims and values cannot be realized

¹ The fact of reasonable pluralism is also mentioned in JFR, pp.4, 9, 33-34, 36, 73, 77 and 84, and IPPR, pp.573 and 576-77. As the remarks made in these two writings do not depart from those made in PL in any substantial way, my main focus will be PL unless otherwise stated.

I am sure that Rawlsian scholars would have welcomed a section of one of his PL lectures devoted to the fact of pluralism. As I have said, we have to make do with his series of discontinuous remarks. For relevant sections, see PL, Lecture I, Section 6.2, and Lecture II, Sections 1-3 as well as his two Introductions (pp.xv-xxxvi and pp.xxxvii-lxii).

² IOC, pp.423-25.

³ Ibid., p.424.

simultaneously. Given this diversity of conflicting and incommensurable conceptions of the good, a political conception must detach itself from any one of those general and comprehensive views. Interestingly, the comprehensive views singled out here are perfectionism, utilitarianism, idealism and Marxism, all of which are at least general (given the scope of their application), and some may even be comprehensive (given the extensiveness of their domain of values).⁴

The importance of the fact of pluralism resides in it being a *permanent* feature of democratic societies rather than merely a historical *contingency* or accident. The level of pluralism experienced in current democratic societies is the result of an ongoing historical process that emerged in the Sixteenth century with the Reformation, where religious pluralism led first to toleration, and then to constitutional government. Once basic rights and liberties were constitutionally protected, an originally limited pluralism, encompassing only different religious confessions, became gradually broader in scope. Pluralism is a permanent fact because it is sanctioned by the political and social logic of historical development of, at least, the last four centuries,⁵ where basic rights and liberties define the conditions under which a diversity of comprehensive views can flourish.

There is one last point in this paper that is only mentioned in passing, but that nonetheless contains an interesting observation. Rawls concedes that the question about the incommensurability of comprehensive doctrines may be difficult to assess. However, the important point is that incommensurability needs to be viewed as a

⁴ *Ibid.*, p.424n.

⁵ I say at least because it may be claimed that pluralism has always been a feature of reasoning. One may see the controversies of Epicureans and stoics in Greek philosophy as instances of pluralism. We may think of other examples such as the polemic between Jansenists and Gallicanists, or the debates about the moral status of the native peoples of America during the Spanish conquest and colonization. These controversies are also driven by an incommensurability of values. We could say that although they may be a product of the burdens of judgement, they are nonetheless a historical contingency or accident. The political and social conditions were not ripe for full pluralism to develop. However, none of this affects Rawls's views of pluralism, or indeed my reconstructed account.

political fact as ‘there is no available political understanding as to how to commensurate these conceptions for settling questions of political justice.’⁶ In other words, the extent of full incommensurability cannot really be decided politically as we may have to invoke comprehensive doctrines to do so. However, a political conception of justice is concerned with the implications of such incommensurability for the purposes of working out a conception of justice suitable to democratic societies. We may thus say that incommensurability is political rather than metaphysical.

In PL Rawls introduces a crucial refinement into his notion of the fact of pluralism. He adopts from Cohen the distinction between simple and reasonable pluralism.⁷ The simple fact of pluralism, Cohen argues, is indiscriminating and entails the logical possibility that conceptions of justice may be adjusted to bring unreasonable views into the overlapping consensus.⁸ An adequate understanding of pluralism must not rely on the fact of diversity as given, but on diversity as arising out of the free exercise of reason, otherwise overlapping consensus could be open to the charge of being a mere compromise driven by interests or by an accommodation to power. The problem, of course, is how to discriminate between those exercises of reason that lead to reasonableness and those that lead to unreasonableness. In Cohen’s terms, arguments that justify the exercise of political power on the basis of norms, values and ideals that may be acceptable to all ‘who are prepared to listen to reason’⁹ fall within the range of reasonable views. By contrast, arguments designed to impose their set of values on others who may not share them fall outside the domain of reasonableness. Rawls borrows Cohen’s conceptual distinction and this allows him to exclude from the overlapping consensus views based on ‘self- and class interests, or

⁶ IOC, p.425n.

⁷ ‘Moral Pluralism’, pp.282-87. See also PL, p.36.

⁸ Ibid., p.284.

⁹ Ibid., p.286.

on peoples' understandable tendency to view the political world from a limited standpoint.'¹⁰ Rawls incorporates this distinction into his theory to the extent that he focuses exclusively on the fact of reasonable pluralism. The upshot of this is that reasonableness sets limits on both the scope and extent of disagreement and, hence, of pluralism.

1.Reasonable Pluralism: Features and Implications for Political Liberalism.

This section contains a general exposition of reasonable pluralism that is in turn divided into two subsections. Firstly, I look into the features of Rawls's account of pluralism and, secondly, I discuss the consequences of his account for the project of political liberalism.

Features of Reasonable Pluralism

The first significant feature of reasonable pluralism is its almost exclusive focus on doctrinal conflict. The fact of reasonable pluralism is defined as 'the fact of a plurality of reasonable but incompatible comprehensive doctrines.'¹¹ A basic feature of modern democracies is the existence of this variety of incompatible (but reasonable) religious, philosophical and moral doctrines. Interestingly, Rawls does not use the adjective 'incommensurable' to refer to such doctrines (as in IOC), instead he points to the doctrines' mutual incompatibility.¹² Trying to establish the incommensurability of two or more comprehensive doctrines may be a controversial exercise in itself, and it may require using arguments from comprehensive doctrines themselves. Mutual incompatibility simply points at the fact of incompatibility itself, indicating how the goals and values of two or more comprehensive doctrines are unlikely to be

¹⁰ PL, p.37.

¹¹ PL, p.xix.

¹² Ibid., p.xviii.

simultaneously realized.¹³ And, to complete the picture, we must add that reasonable pluralism may contain doctrines that are not themselves liberal, as not all free exercises of human reason result in more or less liberal doctrines.¹⁴ Hence, reasonable pluralism refers to the existence of a plurality of incompatible, but reasonable, liberal and non-liberal, religious, philosophical and moral doctrines. And, as we said above, this plurality of reasonable doctrines is not a contingent, but a permanent feature of modern democracies.¹⁵

The second feature of reasonable pluralism is that it arises as a consequence of the burdens of judgement,¹⁶ or the sources and causes of reasonable disagreement. The idea of the burdens of judgement is used to explicate how reasonable disagreement obtains and persists. A crucial point here is that the burdens of judgement are themselves properties of the process of reasoning as such and in no way indicate a possible failure on the part of the reasoning agent.¹⁷ Moreover, they apply to reasoning agents as both reasonable and rational, and they cover the complete process of reasoning from the selection and assessment of the evidence to the evaluative stages. These sources of reasonable disagreement are carefully distinguished from the sources of unreasonable disagreement, which by definition, are incompatible with the reasonableness of those who disagree. This list includes 'prejudice and bias, self- and group interest, blindness and wilfulness', ignorance, perversity and 'rivalries for power, status and economic gain.'¹⁸

¹³Political liberalism assumes that, *for political purposes*, a plurality of reasonable yet incompatible comprehensive doctrines is the normal result of the exercise of human reason within the framework of the free institutions of a constitutional democratic regime'. Ibid., p.xviii (emphasis added).

¹⁴ Ibid., pp.xxxix-li. Of course, non-liberal doctrines do not necessarily have to be illiberal ones.

¹⁵ Ibid., pp.36, 136 and 216f.

¹⁶ Ibid., pp.54-58.

¹⁷ Ibid., p.55.

¹⁸ Ibid., p.58.

Rawls considers two kinds of sources of reasonable disagreement: those that apply generally to theoretical reasoning, and those that specifically apply to reasonable and rational agents in their exercise of practical reason.¹⁹ Amongst the burdens that apply to theoretical uses of reason are:

- the complexity and conflicting character of the empirical and scientific evidence related to a particular matter
- the difficulties therefore of assessing and evaluating evidence
- the relative weight of different considerations involved in a case
- the different types of judgment involved in weighting up the different aspects of a case
- the indeterminacy of moral and political concepts
- the various types of possible judgements and interpretations required by indeterminacy
- the way our particular experiences shape our assessments, evaluations and judgements in different types of cases.

Burdens of judgement that apply to agents conceived as reasonable and rational include:

- the different (and often incompatible) types of normative considerations involved in a specific matter
- the difficulties involved in making an assessment
- the constraints posed by the social and political system under which moral and political values are to be pursued
- the necessity to select and prioritize some at the expense of others.

The significance of this account and the conclusion that Rawls draws from it is that the conditions under which most of our important judgements are made are such that it is difficult, if not impossible, to see how able reasoning agents, using their full powers, can all reach the same conclusion.²⁰ General agreement amongst free

¹⁹ In this section I follow Rawls's account in *Ibid.*, pp.56-57.

²⁰ *Ibid.*, p.58.

reasoning agents is therefore not to be expected. The fact of reasonable pluralism thus is a logical consequence of the burdens of judgement argument.

Although the distinction between simple and reasonable pluralism was mentioned at the outset, we need to say a bit more about the third feature of reasonable pluralism, namely, its focus on reasonableness. One of the problems when looking at the reasonable is the extension of its scope. The reasonable may be a predicate of doctrines, conceptions of justice or persons.²¹ In relation to pluralism, reasonableness applies mainly to doctrines. A reasonable doctrines has the following characteristics:

- it is a more or less coherent exercise of theoretical reason on the main aspects of human life (religious, philosophical and moral)
- it selects, and prioritizes certain values so as to produce a clear and easily understandable view of the world
- it is also an exercise of practical reason as it may need to balance the respective weight of its different values
- it is normally based on generally recognized doctrinal traditions.²²

The account of reasonable doctrines is, Rawls admits, as wide as it can possibly be. Most doctrines are therefore granted the presumption of reasonableness. However, in relation to pluralism, reasonableness may also be a feature of persons. Reasonable persons are defined as those who are willing 'to propose and honor fair terms of cooperation' and 'to recognize the burdens of judgement and to accept their consequences.'²³ A person may be unreasonable when holding a reasonable doctrine in an unreasonable manner.²⁴ For instance, a person may hold a normally comprehensive doctrine –perfectionism, say- but may try to do everything possible to impose such a doctrine on others had she the political power to do so. Such a person,

²¹ All of our references so far have been to the reasonableness of doctrines. For references to the reasonableness of political conceptions, *Ibid.*, pp.xxxviii and 156f, and to the reasonableness of people, *Ibid.*, pp.49f, 54 and 81.

²² *Ibid.*, p.59.

²³ *Ibid.*, p.49n.

²⁴ *Ibid.*, p.60n.

although holding a reasonable doctrine, must be regarded as unreasonable. Hence, Rawls's account of reasonable pluralism, despite its initial focus on doctrines, exhibits a further constraint: holders of comprehensive doctrines must themselves be reasonable.

To sum up what we have said so far, Rawls's account of pluralism has the following significant features:

- it is a product of the burdens of judgement
- it is a permanent feature of modern societies and therefore is neither an accident nor a temporal contingency
- it focuses on doctrinal conflict and on the incompatibility of doctrines
- it is limited by the reasonable.

Consequences of reasonable pluralism

I have attempted to flesh out Rawls's idea of reasonable pluralism. By now it should be clear the centrality of this notion for the project of political liberalism, in particular as a starting point for the method of political constructivism. Before I proceed to assess it, I shall outline Rawls's own considerations of the implications of reasonable pluralism for his project as a whole.

The first consequence of the fact of reasonable pluralism is that given the diversity of reasonable, yet incompatible, comprehensive doctrines, it is impossible, or at least highly unlikely, to obtain political agreement on the truth of comprehensive doctrines.²⁵ The burdens of judgement limit the extent of feasible political agreements. These may be hard to come by, but not as stubborn as political judgements on comprehensive doctrines. In searching for a suitable basis of political agreement,

²⁵ Ibid., p.63.

comprehensive doctrines should be bypassed. Given the practical impossibility of reaching political agreement on comprehensive doctrines, a society can only be united on any one of such doctrines through coercion. Any comprehensive doctrine that aspires to well-order a society marked by the fact of reasonable pluralism can only do so through the use of the machinery of the state. Rawls refers to this as the fact of oppression.²⁶

A second consequence is that both the fact of reasonable pluralism and the fact of oppression lead to the idea of a political conception of justice. The existence of incompatible, yet reasonable, comprehensive doctrines requires that we look for a political conception of justice as the only possible basis of social unity since 'rather than confronting religious and nonliberal doctrines with a comprehensive liberal philosophical doctrine, the thought is to formulate a liberal political conception that those nonliberal doctrines might be able to endorse.'²⁷ This political conception is to serve as a reasonable public basis of justification for fundamental political questions such as constitutional essentials and matters of basic justice.²⁸

A third implication of the fact of reasonable pluralism is that there is a dualism of perspectives between, on the one hand, the point of view of the political conception and, on the other hand, the multiple points of view of comprehensive doctrines. The nature of such a dualism is not philosophical, but political.²⁹ Having a dual perspective entails that citizens are able to apply both points of view to their respective domains with the perspective of the political conception defining the limits of what can be publicly discussed and politically settled. This dualism is not metaphysical, nor does it

²⁶ Ibid., p.37.

²⁷ Ibid., p.xlvii.

²⁸ Ibid., pp.xxi and xxiii.

²⁹ Ibid., p.xxiii.

rest on any particular conception of personhood; it arises as a result of pluralism as a feature of democratic political culture.

The final conclusion drawn from the account of reasonable pluralism to a certain extent complements the previous three. A well-ordered society contains a number of reasonable political conceptions, therefore it nurtures two different types of disagreement, doctrinal and political: in a well-ordered society ‘there is both the fact of reasonable pluralism and a family of reasonable though differing liberal political conceptions.’³⁰ Any liberal political conception that is reasonable (i.e. honours the fair terms of cooperation and recognizes the burdens of judgement) can be a candidate for becoming the focal point in an overlapping consensus.³¹ In a well-ordered society, given the fact of reasonable pluralism, political discussion is dominated by competition among the set of reasonable liberal political conceptions.³² The fact of reasonable pluralism takes us to this (somewhat surprising) conclusion.

2.Problems with Rawls’s account.

I say that this is a surprising conclusion because, in a democratic society, we would expect political discussion to be more inclusive.³³ If political liberalism grants that, *for political purposes*, a variety of incompatible, yet reasonable, comprehensive views are the product of free and equal reasoning agents,³⁴ how does Rawls get from this seemingly inclusive starting point to that reductionist endpoint? If the problem of political liberalism is to work out a political conception of justice that ‘a plurality of reasonable doctrines, *both religious and nonreligious, liberal and nonliberal*, may

³⁰ Ibid., p.xxxviii.

³¹ Ibid., p.xlix.

³² Ibid., p.xiviii.

³³ I am tempted to say *far* more inclusive.

³⁴ Ibid., p.xviii, emphasis added.

freely endorse,³⁵ such conclusion appears less than satisfactory. It does seem to conflict with what Rawls says elsewhere about the circumstances of justice. These circumstances reflect inescapable historical conditions, therefore when working out a suitable political conception for a democratic society, plurality needs to be taken as given. However, even here, plurality is strictly identified with reasonable pluralism.³⁶ The question is whether simple, and not reasonable pluralism, is a better reflection of the circumstances of justice in modern democratic societies.

One problematic aspect of the distinction between simple and reasonable pluralism is this. We concede, following Cohen, that our concern should be with different understandings of value, and not with a variety of narrow and self-interested points of view.³⁷ But such moral pluralism (itself the result of the constitutional protection of basic rights and liberties) is limited even further. The fact of reasonable pluralism implies that, within the set of moral views, there will be some moral understandings of value that 'will be reasonable, and permissibly taken by their adherents to be true.'³⁸ Whereas there is divergence with respect to people's overall moral conceptions (that is, with respect to the whole truth), there is convergence on a subset of the moral truth, so that in an overlapping consensus, supporters of each view within it 'hold that *nothing but the truth*' is endorsed.³⁹ When Rawls takes this distinction from Cohen, he dispenses with the category of truth applied to the focus of an overlapping consensus,⁴⁰ but he relies on the correctness of at least one of the comprehensive doctrines within an overlapping consensus to provide the required

³⁵ Ibid., p.xl, emphasis added.

³⁶ JFR, p.84.

³⁷ 'Moral Pluralism', p.281.

³⁸ Ibid., p.282.

³⁹ Ibid., p.283.

⁴⁰ PL, p.xxii, 94 and 116. This is because 'when we speak of the moral truth of a political conception, we assess it from the point of view of our comprehensive doctrine.' Ibid., p.126. To this we may add: 'and not from its own political viewpoint.'

moral ground and thus to be able to avoid the charge of accommodation.⁴¹ From our partial standpoints, there are different criteria of correctness. However, should one of such criteria be found correct, all parties to the consensus (all reasonable comprehensive doctrines) should be seen as correct, *politically speaking*.⁴² My purpose here is not to elaborate on the Rawlsian notions of truth and correctness, however complex they may be, but to illustrate the artificiality of his distinction between truth, correctness and reasonableness. Let me reproduce one of the most puzzling passages of PL:

‘...if any of those reasonable comprehensive doctrines supports only true moral judgements, the political conception itself is correct, or close thereto, since it is endorsed by a true doctrine. Thus, the truth of any one doctrine in the consensus guarantees that all the reasonable doctrines yield the right conception of political justice, even though they do not do so for the right reasons as specified by the one true doctrine. When citizens differ, not all can be fully correct, for some are correct for the wrong reasons; yet if one of their doctrines should be true, all citizens are correct, politically speaking: that is, they all appeal to a sound political conception of justice. Besides, we always think our own view is not only reasonable by also moral speaking true, or reasonable, as the case may be. Thus, everyone in a reasonable overlapping consensus finds the political conception acceptable, whatever each person’s final criterion of correctness may be.’⁴³

This is a puzzling section mainly because Rawls’s account of reasonable pluralism and, in particular, his idea of the burdens of judgement, should have prevented him from appealing to truth. It is one thing to concede (as he does) that comprehensive doctrines may use that category themselves, but quite another to offer it as further support for overlapping consensus, and hence for the political conception. Note that the claim is not that the political conception is itself true, but rather that in being supported by a true comprehensive doctrine, the political conception is correct. From this it further follows that the set of reasonable doctrines all support the correct

⁴¹ Ibid., pp.126-29.

⁴² Ibid., p.128.

⁴³ Ibid.

political conception, albeit some for the wrong reasons. Let us put this in the form of a deductive argument:

Premise 1: At least one of the reasonable doctrines within overlapping consensus supports true moral judgements.

Premise 2: Because of this, the political conception itself is correct.

Therefore: All reasonable doctrines within an overlapping consensus yield the right conception of political justice.

This is the argument as far as Rawls puts it. But there seems to be something missing here. There are in fact two unstated premises in this argument. Let us work out the argument fully:

Premise 1: At least one of the reasonable doctrines within an overlapping consensus supports true moral judgements.

Unstated premise 1: The truth of at least one reasonable doctrine yields the correctness of the political conception.

Premise 2: The political conception is therefore correct.

Unstated premise 2: If at least one reasonable doctrine yields the political conception, all reasonable doctrines must yield it.

Conclusion: All reasonable doctrines within overlapping consensus yield the right conception of political justice.

The problem is that the moves defined by these two unstated premises are prevented by Rawls himself. The first premise must be rejected for two reasons: 1) because it is reasonableness, rather than truth which confers moral ground to the political conception, and 2) because the favoured political conception must be sound, rather than correct.⁴⁴ The second premise also has to be rejected because it defeats the very object of overlapping consensus, that of providing multiple support for the political conception. In another passage of the same section Rawls says that whether

⁴⁴ The difference between the two may be summed up in one sentence: 'Although there may be various sound political conceptions, only one conception can be correct.'

or not we regard any reasonable doctrine as true is something that falls outside the scope of the political conception itself. This appears to contradict what has been said above. I do not, however, concede any exegetical defeat since the argument presented above is not an argument from within a comprehensive doctrine, but from within political liberalism itself.

So far my aim has been to illustrate some of the internal difficulties faced by Rawls's particular account of pluralism (i.e. reasonable pluralism). I will now try to problematize some of the issues that arise as a result of posing the following questions: Firstly, what is so significant about doctrinal conflict for Rawls to focus almost exclusively on it? Secondly, why do the burdens of judgement lead to reasonable, rather than to simple pluralism? Thirdly, why do the burdens of judgement lead to the conclusion that a suitable conception of justice must be a liberal political conception of justice? And fourthly, is reasonableness a predicate of doctrines, political conceptions or persons? In discussing these aspects, we need to keep in mind the question that political liberalism is trying to address: how and on what grounds are free and equal, reasonable and rational people to uncover the suitable basis of justification of a constitutional regime? I will refer to this as the framing question of political liberalism.

a) Three kinds of conflict

In addressing these questions I do not wish to cast doubt on the importance of pluralism as the starting point of the project of political liberalism. The aim is to elucidate whether, and to what extent, the Rawlsian version of reasonable pluralism is the most adequate for his own purposes. In particular, we need to see whether there are aspects of pluralism that have been sidelined and why. It is worth recalling that

Rawls's initial concerns with pluralism arise as a result of his worries about stability. In modern democratic societies pluralism is part and parcel of the forces that may make a political order unstable. But there is also another, more positive, aspect: pluralism is the result of reasoning under free conditions. It is therefore a democratic fact. Hence, the importance of pluralism lies in it being both a democratic result and starting point. It is, above all, a democratic concept, and not a value derived from a liberal comprehensive doctrine. This is also, I think, Rawls's understanding of the importance of pluralism. But if pluralism is, on the one hand, a source of instability and, on the other hand, a democratic fact, it is difficult to see how an account that focuses on doctrinal conflict is at all satisfactory. Two reasons present themselves here: 1) it is quite likely that there may be other sources of conflict (beyond doctrinal conflict) that may threaten stability, and 2) given our framing question, there may be other sources of disagreement that may be more relevant than doctrinal disagreement.

To be fair, Rawls does acknowledge that modern societies normally contain three kinds of conflict: doctrinal (between comprehensive doctrines), theoretical (resulting from the burdens of judgement), and sociological (the product of social diversity and difference). However, he deals with them in a strange kind of way. Political liberalism can only mitigate, but not eliminate, doctrinal conflict. Sociological conflict, by contrast, can be fully reconciled in a just constitutional regime and settled by a political conception of justice. Finally, theoretical conflict will remain and affect the possibilities for political agreement.⁴⁵ Leaving aside theoretical conflict which, in a sense, supersedes the limits of normative theory, it seems rather ingenuous to think that doctrinal and sociological conflict can be kept fully separate for the purposes of working out the most adequate conception of justice for a

⁴⁵ *Ib.d.*, p.lx and IPPR, p.613.

democratic society. Hence, we need to clarify why, in Rawls's view, doctrinal conflict is so significant. The explanation could be found in the views of Rawls the historian. The origins of political liberalism as a project, Rawls claims, can be traced back to post-Reformation society and its politics. The diversity of antagonistic religious doctrines necessitated a political settlement in which different creeds could coexist more or less harmoniously. This *modus vivendi* gave rise to the principle of religious pluralism and to 'something like the modern understanding of liberty of conscience and freedom of thought.'⁴⁶ This is the historical precedent that shows how stable pluralist societies are possible. This historical understanding of his project leads him to formulate the problem of political liberalism in a particular way: 'How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable religious, philosophical, and moral doctrine?'⁴⁷ All he does therefore is to open up the scope of reasonable disagreement so as to include the different types of doctrines that normally flourish in modern societies.

But, with regards to pluralism, modern societies differ in important ways from post-Reformation society. The relevant disagreements in the latter are about the good, in particular about the religious good. The problem of political justice in post-Reformation society was that particular religious doctrines greatly determined the kind of political authority that their followers were to submit to. A *modus vivendi* was the only alternative to war. But even then, there were political conflicts within comprehensive doctrines themselves. Take, for instance, the history of seventeenth century France which was marked by the conflict between Gallicanism and Ultramontaniam. Gallicanists maintained that the power of the papacy should be

⁴⁶ PL, p.xxvi.

⁴⁷ Ibid., p.xxvii.

limited by the temporal powers of monarchs. The doctrine of Ultramontanism⁴⁸ argued for the superiority of papal authority over states and monarchs. Hence, even within Catholicism there were different understandings of the extent and limits of both secular and religious authorities. This suggests that, even within comprehensive doctrines, there may be disagreements about matters bearing on issues of justice. Such disagreements might have been doctrinal in their origins, but at some point they have ceased to be so. This example has only limited value. Political liberalism would suggest that this example shows why irreconcilable disagreements required a *modus vivendi*, and Gallicanism had implicitly acknowledged this by submitting to an appropriate (secular) political authority. Ultramontanism, on the other hand, in rejecting the legitimacy of state power, may be classed as an unreasonable doctrine. However, other examples of disagreement within a comprehensive doctrine do not work so nicely in the case of modern societies. Think, for example, of Marxism and Catholicism, two intrinsically incommensurable doctrines that are on opposite poles.⁴⁹ Are they, however, politically incompatible to the same extent as they are metaphysically incommensurable? If we now think of those within Catholicism that follow Liberation theology and, in particular, the doctrine of the ‘preferential option for the poor’, we may conclude that, in an extrinsic political sense, they are quite close to Marxism, and quite distant from, say, libertarianism. This kind of conflict seems more relevant to a political project, and to a normative theory of politics, than doctrinal conflict *per se*.

A possible reason of Rawls’s focus on doctrinal conflict besides the historical explanation just mentioned is this. Political liberalism, he claims, can only mitigate,

⁴⁸ Ultramontanism etymologically means ‘beyond the mountains’ where mountains referred to the Alps; this, of course, meant Rome.

⁴⁹ They are incommensurable to the extent that both would regard the other as such from within its own comprehensive perspective. This should be sufficiently clear by a mere mentioning of, on the one hand, the theory of historical materialism and, on the other hand, the theology of the history of salvation.

but not eliminate, the conflict between comprehensive doctrines because they 'are, politically speaking, irreconcilable.'⁵⁰ We can think of the problem as one of political incompatibility. Comprehensive doctrines are likely to make claims on the state where such claims are largely incompatible with one another, and the state cannot side with any one comprehensive doctrine if the fact of oppression is to be avoided. For instance, Marxism and libertarianism are two of such comprehensive views and they are, therefore, bound to make claims on the state that are irreconcilable. The state must remain, in a sense, equidistant with respect to these two (and the whole set of) comprehensive doctrines.⁵¹ But, as we said above, democratic societies are characterized by a second kind of conflict, a conflict that we termed 'sociological'. This kind of conflict derives from people's 'different status, class position, and occupation, or from their ethnicity, gender, and race.'⁵² Rawls is rather optimistic about the success of political liberalism when dealing with this. But take, for example, conflict based on gender. The acknowledgment and subsequent political articulation of this type of conflict has given rise to the elaboration of different political views, some of which may fall within the set of comprehensive doctrines (radical feminism, for instance). The question is thus: how can political liberalism claim to reconcile us to gender conflict if it is to remain equidistant between radical feminism and, say, patriarchal views? It seems that some comprehensive views may not just be unavoidable in political argument, but may be required in order to communicate the force of political injustices. It may be said here that *the proviso* does allow for the introduction of comprehensive doctrines into political discussions conducted on the

⁵⁰ Ibid., p.lx.

⁵¹ The term 'equidistant' is, I think, more appropriate than the term 'neutral'. As Rawls explains, political liberalism is only committed to neutrality in the sense that 'the state is to secure equal opportunity to advance any permissible conception', and 'the state is not to do anything intended to favor or promote any particular comprehensive doctrine rather than another, or to give greater assistance to those who pursue it.' Ibid., p.193.

⁵² Ibid., p.lx.

basis of public reason.⁵³ However, this is an argument made from within a liberal political conception of justice, or from a set of liberal conceptions. Here we are not considering how citizens are to conduct their political discussions in a constitutional democracy. We are rather looking at the starting point, that is, at what may be the most suitable basis of peaceful coexistence amongst people who profoundly disagree on a whole host of matters, and, in relation to this, we are also asking whether Rawls has provided an adequate account of relevant disagreement. This section has expressed some doubts about his account of reasonable pluralism, but the critical points just made require further elaboration.

b) Reasonable pluralism and the burdens of judgement

This section is concerned mainly with two questions. The first question is whether political liberalism deploys the predicate 'reasonable' in a clear and consistent manner. This is an important aspect because it may have implications for the internal coherence of the project. The second question is even more crucial: it asks whether the account of the burdens of judgement may lead to simple, as opposed to reasonable, pluralism.

Political liberalism uses the predicate 'reasonable' in a threefold way: it applies to comprehensive doctrines, conceptions of justice and persons. The reasonable is closely associated with the idea of the burdens of judgement since, as we have seen, accepting and honouring the burdens of judgement is a necessary condition of being reasonable. However, different criteria of reasonableness hold depending on the particular subject for that predicate. The criteria of reasonableness are much looser when they apply to comprehensive doctrines than when they apply to the political

⁵³ IPPR, pp.591-94.

conception or to persons. In fact, most doctrines are generally presumed to be reasonable unless otherwise shown. By contrast, reasonableness is considerably more stringent as a predicate of political conceptions or persons, for there are two necessary conditions attached here: the acceptance of the burdens of judgement and the willingness to propose and honour fair terms of cooperation. Hence, contrary to the views of some commentators, reasonable doctrines are not only those that accept the principles of political liberalism. This will certainly be self-defeating.⁵⁴ But if this is the case then, on what basis does Rawls label certain views as unreasonable given the inclusiveness of his definition of reasonable comprehensive doctrines? To be sure, he does not want to rule many things out. Unreasonableness is mainly a feature of persons. It is persons who decide to hold reasonable doctrines in an unreasonable manner by rejecting the fact of oppression. So, we might ask Rawls: is, say, racism unreasonable in itself, or is it unreasonable because racists themselves are unreasonable? His likely answer will be that whereas racists are certainly unreasonable, racism does not even qualify as a viable comprehensive doctrine. Hence, there are three criteria of reasonableness/unreasonableness:

1. as applied to doctrines, most viable doctrines (i.e. long-standing traditions of thought and practice that have stood the test of time) are reasonable unless proven otherwise.⁵⁵
2. as applied to conceptions of justice, reasonableness requires the acceptance of both the burdens of judgement and of fair terms of cooperation.⁵⁶

⁵⁴ See, for instance, Moller Okin, S., 'Political Liberalism, Justice and Gender' and Baumeister, A.T., *Liberalism and the 'Politics of Difference'* (Edinburgh: Edinburgh University Press, 2000), esp.ch 3. For Rawls, however, unreasonable doctrines are those 'that reject one or more democratic freedoms.' PL, p.64n. Also, Mulhall and Swift suggest that reasonableness entails the acceptance of the political conception of the person and of society that are the basis of political liberalism. See 'Rawls and Communitarianism' in *The Cambridge Companion to Rawls*, pp.460-87, p.482.

⁵⁵ PL, p.59.

⁵⁶ *Ibid.*, pp. xlix and 49n.

3. as applied to persons, reasonableness requires the same as in 2 plus the acceptance of the fact of oppression.⁵⁷

Recall, however, that our interest in the reasonable arose in the first place because of its association with pluralism. We also showed at the outset that pluralism was the logical consequence of the burdens of judgement argument. Hence, unless there is a technical failure in the application of the burdens of judgement or unless the reasoning agent is herself unreasonable, most exercises of reason must be classed as reasonable ones. The burdens of judgement would appear to push in the direction of simple, rather than reasonable pluralism. To be more precise, within the term 'reasonable pluralism', the 'reasonable' refers to the reasonableness/unreasonableness of the reasoning agent, and not to the content of particular views. This implies that restrictions on pluralism do not come from the range of views held, but from the features of agents themselves. This is, although not Rawls's, a Rawlsian conclusion.

Part of the problem might have derived from an over-use of the 'reasonable' predicate. Some critics have suggested that Rawls ought to have focused on reasonable persons rather than reasonable doctrines.⁵⁸ However, he does not focus on reasonable persons and doctrines simultaneously. We need to distinguish here between the first and the second stages of political liberalism. Reasonable doctrines play a crucial part mainly at the second stage as the parties to the overlapping consensus. By contrast, the focus on reasonable persons is present in both stages as initially demarcating pluralism (first stage) and then, subsequently, the range of comprehensive doctrines (second stage) since 'reasonable persons will think it unreasonable to use political power, should they possess it, to repress comprehensive views that are not

⁵⁷ Ibid., pp.50, 60 and 61.

⁵⁸ For instance, Barry, 'The Search for Stability', pp.898-99, and Jones, P., 'Two Conceptions of Liberalism, Two Conceptions of Justice', *British Journal of Political Science*, 25(1995), pp.515-550, in particular pp.226-7.

unreasonable, though different from their own.’⁵⁹ Hence, what I am advocating here is that, amidst all the confusion created by the over-usage of the term ‘reasonable’, it is possible to interpret the fact of reasonable pluralism in a more extensive way. Remember that here we are concerned with reasonable pluralism as a democratic fact and as the starting point of a political project.

The aim of political liberalism (its framing question) is to investigate how free and equal, reasonable and rational people can uncover the suitable basis of justification on political fundamentals. Pluralism presents the crucial challenge to this key political question. Hence, we have asked whether Rawls’s account of pluralism was adequate for the project of political liberalism, or whether there was something relevant somehow missing in his account. We have tried to show that, by concentrating on doctrinal conflict, Rawls was looking in the wrong place. His account is, therefore, too narrow. Moreover, by not paying sufficient attention to political pluralism, a key aspect of his political project, his account is not adequately focused. However, we have seen that the burdens of judgement seem to collide with his almost exclusive doctrinal focus, but that there are resources within his theory for opening up the basis of pluralism. Hence, reasonable pluralism needs to be understood not as the pluralism that obtains from comprehensive doctrines, but from reasonable persons. Both the burdens of judgement and the requirements of Rawls’s political project demand this more inclusive understanding.

Rawls starts by asking how reasonable disagreement might come about. His answer, as we have seen, is that reasonable disagreement is a result of the burdens of judgement and of their consequences for the exercise of reason under free conditions. This leads to the recognition of the importance of reasonable pluralism. The second

⁵⁹ PL, p.60.

question that we have asked, but that Rawls does not fully ask, is about the kind of pluralism that is relevant for our framing question. Our suggestion has been that the pluralism on which political liberalism ought to focus is a political pluralism of reasonable persons. According to this view, there is no point in asking whether Marxism or libertarianism are reasonable views. We should ask instead whether Marxists and libertarians have a part to play at the starting point of the project, as part of this political pluralism that the burdens of judgement define. Our answer is affirmative, always provided that the advocates of such views are themselves reasonable. There is still one problem remaining, and it is a pressing one. Rawls thinks that the burdens of judgement apply to comprehensive doctrines in particular.⁶⁰ Political agreement, he says, is difficult to obtain, but it is even more difficult to obtain political agreement on comprehensive doctrines. Some critics have suggested that Rawls has over-emphasized the importance of disagreement in his theory.⁶¹ They claim that just because disagreement is expected, this does not mean that it is inevitable, and that the very possibility of reasonable disagreement also entails the possibility of reasonable agreement. This appears somewhat to beg the question. Rawls is rightly interested in disagreement as a democratic fact, and the comprehensive complexity of modern democracies seem to point, at least for the time being, to the existence of widespread disagreement. However, it may well be the case that agreement on comprehensive doctrines is far harder to obtain than political agreement, but this does very little to dispel the objection that the burdens of judgement might apply just as decisively to political conceptions of justice.⁶² I turn to this question in the final part of this section.

⁶⁰ Ibid., p.63.

⁶¹ *Liberals and Communitarians*, p.235.

⁶² Amongst those who consider some of the difficulties of justice as fairness given the burdens of judgement are Copp, D., 'Pluralism and Stability in Political Theory', *The Journal of Political*

c) the burdens of judgement and political liberalism

In the previous section we focused mainly on reasonable doctrines and reasonable persons. However, 'reasonable', as we saw, also applies to conceptions of justice. Since political liberalism dispenses with the category of truth, it considers the political conception as reasonable. A well-ordered society contains a number of reasonable political conceptions. In fact, a well-ordered society is characterized not just by the fact of reasonable pluralism, but also by the existence of a family of reasonable though differing liberal political conceptions.⁶³ Rawls's claim here is that an adequate conception of justice for a democratic society must be reasonable, political and liberal, and these requirements stem from the democratic fact of reasonable pluralism and from the burdens of judgement. We have already looked at the requirements of the reasonableness of a political conception, i.e., the acceptance of the burdens of judgement and of fair terms of cooperation. We also know that a conception of justice is political if it is worked out to apply to the basic structure of a society, if it is presented independently of particular comprehensive doctrines, and if its values are ones that are implicit in the public culture of society.⁶⁴ Finally, a conception of justice is liberal if it specifies various rights, liberties and opportunities, it prioritizes some of those particularly over claims based on the general good and/or perfectionist values, and it proposes some adequate means so that citizens can make effective use of the specified liberties and opportunities.⁶⁵ The question is, given the burdens of judgement, on what basis can Rawls claim that a liberal political conception is the most adequate to well-order a democratic society? Let us look at two related difficulties.

Philosophy, 4 (1996), pp.191-206, Caney, S., 'Anti-perfectionism and Rawlsian Liberalism', *Political Studies*, 43 (1995), pp.248-64 and Mulhall and Swift, *Liberals and Communitarians*, p234.

⁶³ PL, p.xxxviii.

⁶⁴ *Ibid.*, pp.11-15, and IPPR, p.584.

⁶⁵ PL., p.6.

Firstly, there appears to be a kind of circularity at work here. For instance, reasonable disagreement, as we have seen, is sometimes defined as disagreement between reasonable persons⁶⁶ and this involves an account of the burdens of judgement as the very sources of disagreement. However, reasonable persons are those who, among other things, accept the burdens of judgement. Hence, although we can say that the reasonable and the burdens of judgement are mutually supportive, there are no independent grounds for establishing the reasonableness of persons and/or political conceptions. But the criteria of reasonableness are far too stringent. Recall that in chapter 2 we suggested that the burdens of judgement were a crucial aspect of the problem of utopianism. Therefore it is easy to see how the very idea of reasonableness may be undermined by the burdens of judgement themselves. Therefore, Rawls needs to provide a separate account of the reasonableness of persons and of political conceptions.

Secondly, Rawls's definition of a liberal conception seems far too specific. It is roughly a definition of justice as fairness suitably devoid of all particularities rather than of a broad liberal conception. It is, at the very least, a general definition of an egalitarian type of liberalism. But a family of liberal conceptions may include versions that are closer to the libertarian variety than to justice as fairness.⁶⁷ To be sure, there is a kind of tension throughout PL created by the fact that he has two concerns: a defence of his project of political liberalism and also of justice as fairness in particular. The project of political liberalism requires a much looser specification of liberal conceptions. But, even though Rawls acknowledges this problem when he says that

⁶⁶ Ibid., p.55.

⁶⁷ For instance, versions that stop at constitutional consensus. Frank Michelman has suggested that the importance of constitutional consensus is that it democratically secures a libertarian core of basic rights and liberties and that 'any possible conception of justice that does not include [such] libertarian core is beyond the pale of constitutional democracy as Rawls conceives of it.' 'Rawls on Constitutionalism and Constitutional Law, *The Cambridge Companion to Rawls*, p.409.

‘the most reasonable political conception of justice for a democratic regime will be, *broadly speaking*, liberal,⁶⁸ he still requires that general liberal conceptions should protect basic rights, assign them a special priority and insure that citizens possess the necessary materials means to effectively use the basic rights.⁶⁹ Some liberal conceptions would stop at the second requirement and forgo the third one. But, the crucial point is this. It may well be the case that the fact of pluralism requires a reasonable liberal political conception to settle the problems of stability and justification in modern democracies, but the burdens of judgment argument cannot sanction this. In fact, the burdens of judgement push the whole project of political liberalism into a kind of quasi-scepticism (something that Rawls explicitly rejects), although not for the reasons that he contemplates.⁷⁰ The problem of scepticism that I am considering here is not whether abstracting from comprehensive doctrines implies an indifference to truth,⁷¹ but whether he succeeds in shielding the family of liberal political conceptions against the burdens of judgement.⁷² Given that these burdens ‘lead us to recognize that there are different and incompatible liberal political conceptions’⁷³, why aren’t these be exposed to the burdens of judgements to the same extent that comprehensive doctrines are exposed? Is there anything in the definition of these burdens that prevents their application to liberal political conceptions?

As an illustration of Rawls’s problems, let us take an example of the burdens of judgement that applies to agents regarded as reasonable and rational. Take, for instance, the burden that states that ‘any system of social institutions is limited in the values it can admit so that some selection must be made from the full range of moral

⁶⁸ Ibid., p.156, emphasis added.

⁶⁹ Ibid., pp.156-7. See also IPPR, pp.581-2.

⁷⁰ PL, pp.150-54.

⁷¹ This is Rawls’s understanding and his answer in Lecture VI, Section 4 is designed to rebut this charge.

⁷² Similar problems are discussed by Mulhall and Swift in ‘Rawls and Communitarianism’, pp.481-485.

⁷³ PL, p.xlix.

and political values that might be realized.⁷⁴ There is no doubt that a particular polity can only realize some values at the expense of others if only for the straightforward reason that certain sets of values are likely to be incompatible with others. But it is not merely a selection of values like freedom and equality that is at stake here. Their different interpretation, priority and ordering are also important. Even though the burdens of judgement may warrant the selection of freedom and equality (and even this much may be questionable), it is difficult to see how we can get from the burdens of judgement to particular interpretations of those values. Note that the point is not that freedom and equality cannot be selected, but rather that the selection cannot be guaranteed. To the three types of conflict that Rawls acknowledges, we need to add a fourth one, conflicts between political values and their interpretations. A possible explanation of the omission of this type of conflict may be that Rawls has assumed that particular interpretations of political values are always the product of their embeddedness in comprehensive doctrines. It is as if political conflict and the values that motivate it could not function if independent from wider realms of values. But he does acknowledge the possibility of this independence when he discusses the pluralist view.⁷⁵ However, the pluralist view only surfaces at the second stage of overlapping consensus, once a liberal political conception (or a family thereof) has been selected and it is not as such part of the focus of initial conflict.

The conclusion to be extracted from this is that the burdens of judgement are heavily overloaded. Whereas they work as an account of both the sources of disagreement and the limits of possible agreement, they cannot at the same time supply one of the crucial premises of reasonableness, and neither can they sanction the selection of a liberal political conception as the most adequate conception of justice

⁷⁴ Ibid., p.57.

⁷⁵ Ibid., p.170.

for a democratic society or at least, not unless their status is significantly revised. By this I mean that it may be possible after all to interpret the burdens of judgement as a strictly political notion. According to this reading, the burdens of judgement are the sources that account for the democratic fact of pluralism, that is, for the existence of widespread disagreement at different levels, and this includes interests, theories and their application, and values and their selection and ordering. This interpretation also concedes that people may not accept the burdens of judgement other than for political purposes. For instance, a Marxist may regard political and ideological disagreement as the result of false consciousness on the part of non-Marxists. But this would not constitute by itself a sign of unreasonableness, for, as we have suggested, reasonableness and unreasonableness fall on the person. It is persons, and not doctrines, ideologies or views that are reasonable or unreasonable. And finally, as the reasonableness of persons cannot be defined by their acceptance of the burdens of judgement, we are left with a much thinner conception of the reasonable person, a view that is more appropriate for a political project such as Rawls's. Hence, we start with pluralism as a democratic fact (which is accounted for by the burdens of judgement), and we recognize that pluralism sustains political disagreement. If the aim is then to reach some kind of political agreement so that at least some fundamental questions can be settled, we can only really assign two conditions to the reasonable person: 1) the willingness to cooperate and to abide by the terms of cooperation, and 2) the acceptance of the fact of oppression. Anything more would not get us round the problems identified here, in particular it would cut against the democratic fact of pluralism. But anything less would be self-defeating given the original impetus of the project.

Having presented this survey of Rawls's account of reasonable pluralism, I now turn to the key idea of constitutional consensus. The main purpose of this incursion into constitutional consensus is to examine whether, and to what extent, it can provide stability of the required kind.

CHAPTER 5

CONSENSUS AND STABILITY

The last three chapters have been given over to analysing the key ideas of overlapping consensus, stability and pluralism. Before proceeding any further I wish to sum up the preliminary conclusions of my study up to this point. I argued in chapter 2 that the aims of stability and justification pushed the idea of overlapping consensus in two different directions. If overlapping consensus was mainly intended as a solution to the stability problem, Rawls's theory appears somewhat utopian. If, on the other hand, overlapping consensus was conceived principally as a justificatory device, the problem of stability still remains in need of a solution. Hence, if my exposition there is sound, we have to conclude that the identified problem of utopianism calls into question the very possibilities of realization of justice as fairness. These problems led us to examine the question of stability in chapter 3. Initially this was driven by an intuitive appreciation that the problem of stability was far more complex than Rawls had admitted and that the notion of stability was, in a sense, polysemic. Thus, the same term was used to imply separate referents. In a sense the purpose of that chapter was to provide an analytical account of stability and its uses in a way that was sensitive to Rawls's own perception of the changing nature of his theory. My survey concluded by isolating three aspects that are central to the project of political liberalism: institutional stability, political stability and legitimacy. Finally, chapter 4 presented an internal critique of reasonable pluralism and offered an alternative account—one that is more suitable to both the features of democratic societies and of political theorizing in this type of societies. I am now in a position to start making the

first alteration into Rawls's own argument of how justice as fairness is possible. This task is taken up in the present chapter, where I hope to show that Rawls's idea of constitutional consensus, suitably interpreted, already affords a substantial degree of stability of the appropriate kind. This chapter is divided into three parts. The first part is devoted to discussing some preliminary matters surrounding the way in which political liberalism deploys the idea of constitutional consensus. I then turn to an examination of its features, paying careful attention to the differences between constitutional consensus and *modus vivendi*. The final part offers an assessment of this concept in relation to stability. It is in this final part where my alteration to Rawls's argument will be more explicitly effected. In a sense, this chapter is both narrative and analytical. In a very important respect, it sets out to provide a coherent account of constitutional consensus with regards to 1)the idea itself, 2)its content and features, and 3)its accomplishment from *modus vivendi*. But in doing so, I also aim to scrutinize three substantive questions: 1)why constitutional consensus is necessary for the right kind of stability, 2)why it is not yet sufficient, and 3)why a *modus vivendi* falls short of stability of that required kind.

1. Constitutional consensus: Preliminary Remarks.

To be sure, in Rawls's own mind, an overlapping consensus on justice as fairness does seem to present itself as a realistic possibility. In his view an overlapping consensus appears to possess the qualities that make it a realistic utopia. He is nonetheless faced with the task of providing convincing reasons as to why this is the case. It is in this context that constitutional consensus comes to play a part in political liberalism. However, given the importance that Rawls concedes to the need to secure acceptable levels of stability within a democratic polity, to even suggest that

overlapping consensus may be utopian entails in practice admitting to at least the possibility that he might not be able to secure adequately the stability of justice as fairness. Hence, towards the end of Lecture IV Rawls adds an important element to his theory.¹ The suggestion there is that we might be able to make more sense of how justice as fairness can become the focus of an overlapping consensus if we consider a constitutional consensus as a crucial stage in a process that starts with a *modus vivendi* and concludes with an overlapping consensus.² Therefore the idea of constitutional consensus is, from its very conception, associated with the question of the actual prospects for overlapping consensus. A constitutional consensus is understood as a required developmental stage in the transition from a *modus vivendi* to an overlapping consensus. However, this process always looks at overlapping consensus as its goal. But just to clarify possible misgivings: saying that overlapping consensus is the goal of that process does not mean that we should understand such a process in teleological terms. This is not what Rawls intends since he leaves open the possibility that an overlapping consensus might not in the end obtain.³ This is merely an account of a possibility, although in his view, of a strong one.

Now, in accounting for this possibility Rawls considers, what we may call, two 'worst-case scenarios'. Justice as fairness may in the end fail to become the focus of an overlapping consensus for one of two possible reasons: 1) it may fail to become the focus of any kind of consensus beyond a *modus vivendi*, or 2) the obtained consensus (beyond a *modus vivendi*) on justice as fairness may not be stable enough. As Rawls puts it, it may well turn out that '*there are not sufficient political, social, or*

¹ PL, Lecture IV.6.

² See in particular sections 6 and 7 of Lecture IV. Rawls takes the idea of constitutional consensus from Baier's discussion of PnM and IOC in 'Justice and the Aims of Political Philosophy'.

³ PL, pp.xlvii-xlviii.

psychological forces that are able in practice to push for an overlapping consensus.⁴

This deserves careful attention. Without going into too much detail at this point, by the terms in italics I take Rawls to be referring, broadly speaking, to the active players within a democratic society, that is, to its main institutions, political parties, organizations, and to its citizens generally. It is curious to note the terms of definition of the second of the two 'worst-case scenarios'. Here Rawls considers a situation where an overlapping consensus does obtain, but it is one that is not sufficiently stable. This appears somewhat puzzling since an overlapping consensus is, by its very definition, stable. An unstable overlapping consensus is a contradiction in terms. So does this imply that we are left with only the first of the two 'worst-case scenarios'? Let us look at this question from a slightly different angle. A fully-fledged overlapping consensus is Rawls's *complete* answer to the problem of stability. But the question is this: if, at some point, an overlapping consensus begins to appear an unrealistic option, is there anything else that can afford a certain degree of stability (given that a *modus vivendi* is unable to do so) to justice as fairness? The idea is, I think, that we may at some point have a consensus of sorts, and then, as its stability increases, this initial consensus might eventually turn into an overlapping one. Thus the problem of utopianism is not really a single problem, but two: 1) justice as fairness is endorsed as a result of a compromise and is therefore unstable, and 2) justice as fairness becomes the focus of a consensus that falls short of an overlapping consensus. This second case is a very interesting one. If, on the one hand, this less demanding consensus is not completely satisfactory, on the other hand, it allows Rawls to remain positive with respect to the actual possibility of overlapping consensus itself.

⁴ Ibid., p.158.

Let me expand on this. To be sure, the idea of overlapping consensus pre-dates PL, and so does Rawls's strategy of separating questions of justification from questions of stability.⁵ By contrast, the idea of constitutional consensus appears for the first time in PL and, as we have seen, it is primarily intended to deal with some of the practical difficulties highlighted above. Rawls appears to have realized that he needs to provide a more detailed account of the move from a *modus vivendi* to an overlapping consensus, filling in the gaps left in previous, and more schematic, outlines.⁶

Rawls regards an overlapping consensus on justice as fairness as a necessary condition of its stability. A *modus vivendi*, which is, by definition, an agreement based on the need to compromise in the face of political unrest, would not secure that stability. A Rawlsian *modus vivendi* is unstable by definition.⁷ But why? A first apparent reason for this inherent instability is that if a *modus vivendi* is only the result of convergent self-interests or of historical accommodation, any changes with respect to the political and/or historical circumstances within a society could at least question the desirability of previous arrangements, when not throw them into complete disarray. In short, a *modus vivendi* may be able to contain coercion, but it cannot eliminate it altogether. Not, in any case, in a principled way.

Looking at *modus vivendi* a little closer, we can envisage sets of circumstances where it could prove to be the most stable, or indeed the only stable, means of peaceful coexistence; for instance, in the case of extremely divided societies where long-lasting and seemingly intractable political conflicts are a constant feature.

⁵ There is no contradiction in claiming that, although these two set of questions need to be addressed separately, the notion of an overlapping consensus performs a justificatory function as well as a stabilizing one. This is made clear in the 'Reply to Habermas', *Ibid.*, pp.385-95.

⁶ Like the one provided in IOC.

⁷ I think that this is clearly the case although, of course, there may be other valid definitions of *modus vivendi* that are not so unstable at the outset. An example of this is given by Charles Larmore in his *Patters of Moral Complexity* (Cambridge: Cambridge University Press, 1987)

In such cases, the adoption of a minimal agreement along the lines of a *modus vivendi*, without the necessity to dig deep into its moral pedigree, would at first sight appear a better solution from the standpoint of stability.⁸ But this claim seems to misinterpret Rawls's purposes. His concern is not with stability as such, but with the extent to which a liberal conception of justice (and his conception in particular) can be stabilized along democratic lines. The question is not whether a *modus vivendi* would secure the stability of a polity (this is certainly not beyond the realms of probability), but how able a liberal conception of justice (and, again, his conception in particular) is to carry out the task of interpreting, prioritising and ordering the main elements of a constitutional democracy. In a sense, Rawls undertakes, and asks citizens to undertake an exercise of political pedagogy. The purpose of the exercise is to enable us, as both philosophers and citizens, to think reflectively about the basic elements of a constitutional democracy, and about the ways in which justice as fairness would define, articulate and, if necessary, review such basic elements. These elements are, of course, the constitutional essentials and the matters of basic justice.⁹

Before I examine constitutional consensus in some detail, let me just point to a problem that, as we saw in chapter 2, has informed numerous criticisms of Rawls's theory, and that will enable us to put together the themes of this and the next chapter. The problem is the following:

1)if Rawls is even prepared to admit that an overlapping consensus may not be a realizable goal, and

2)if justice as fairness has already been justified as freestanding, then

⁸ But not just in the case of extremely divided societies, for instance, Larmore, whose project is quite close in spirit, if not letter, to Rawls's, argues that in societies, like democratic societies, where people have different ultimate commitments, only a *modus vivendi* can adequately solve stability because it is the only possible way of effecting the required separation between the citizen and the person, see *ibid.*, p.125.

⁹ These are more extensively discussed in Lecture VI rather than in Lecture IV.

3)does a possible failure to solve stability bear in any way on the justification of justice as fairness?

Rawls himself provides an answer to this when he says: 'PL makes no attempt to prove, or to show, that such a[n overlapping] consensus would eventually form around a reasonable political conception of justice. The most it does is to present a freestanding liberal political conception that does not oppose comprehensive doctrines on their own ground and does not preclude the possibility of an overlapping consensus for the right reasons.'¹⁰ Hence, Rawls himself acknowledges that the attainment of an overlapping consensus may be, in a certain sense, contingent on a variety of social and political factors. All that political liberalism needs to do is to present a fully politically justified conception of justice, and then to present a plausible argument for both its stability and its general endorsement by the citizens of a democratic society. The implicit objection that we have been suggesting throughout this thesis is that, precisely because of the contingent character of overlapping consensus, the argument for stability needs to be made rather more watertight. And relatedly, we have been defending that freestanding political justification needs to be protected more adequately so that it does not suffer as a result of such contingencies.

But, the subject-matter of this chapter is not overlapping, but constitutional consensus. This idea, although important to the project of political liberalism, is one of the most underdeveloped within Rawls's latter work and, therefore, before we can assess it in relation to stability, we need to provide here a fuller account of this notion.

A last preliminary matter that I will consider concerns the relation between the two stages of political liberalism. There is something here that troubled me greatly and that created some difficulties of interpretation at the outset. As we know, the first

¹⁰ PL, pp.xlvii-xlviii.

stage concerns the presentation and justification of justice as fairness as freestanding, whilst the second deals with questions of stability.¹¹ However, it appeared rather clear to me that I had to take into consideration the fact that Rawls introduces constitutional consensus at the second stage, when discussing the stabilizing potential of justice as fairness. Because of this, I was led to think that constitutional consensus was part of the second stage, and that therefore the sequential order of political liberalism was the following: we start with freestanding, then we try to see how a politically justified theory can become the focus of a constitutional consensus and, finally, how it can bring about an overlapping consensus. However, I soon realized that justice as fairness was too thick for it to have a prominent place within the constitutional consensus. Rawls makes two points that somehow clarify these difficulties. One is that constitutional essentials should be settled on the basis of reasons specified by one reasonable political conception amongst the family of reasonable liberal conceptions of justice¹² with justice as fairness therefore being only one amongst such political conceptions.¹³ The other point is that the definition of constitutional essentials only covers Rawls's first principle. Thus the process should not be understood as one where we move from a merely political and unreflected consensus (constitutional consensus) to a more thoroughly moral consensus (overlapping consensus) on justice as fairness. But did this imply that constitutional consensus was prior to the freestanding stage? This interpretation had its own difficulties for, if this were the case, Rawls could not grant justice as fairness, or any other political conception from the family of liberal conceptions, any special place in a constitutional consensus. They simply would not have been articulated in any substantial detail. From this it

¹¹ As I suggested in note 5 above, we also know that questions of stability lead to questions of justification. The 'Reply to Habermas' makes this rather clear.

¹² PL, pp.xlix-l.

¹³ Although Rawls clearly views justice as fairness as the most reasonable; however, as he concedes, a view is not an argument. Ibid., pp.xlviii-xlix.

follows that we need to see the two processes as coexistent, although not necessarily as synchronic. For, as I will suggest below, it is a mistake to see constitutional consensus as something fixed; something that took place only once at a particular point in the history of a democratic society. Above all, constitutional consensus is a feature of democracy, rather than a feature of a liberal political conception. From this it follows that we need to detach our examination of constitutional consensus from the internal two-stage division of PL. Constitutional consensus should be addressed on its own as belonging to the realm of democratic theory, and not as such of liberal political theorizing. Constitutional consensus is, to a large extent, independent from justice as fairness.

This last remark deserves some clarification. I illustrate this by reference to what I regard as the underlying story of Lecture IV of PL. Take, what we may call, the two pillars of political liberalism: 1) consideration of the central elements of a constitutional democracy, and 2) the idea of a political conception of justice. The central elements of a constitutional democracy are the constitutional essentials and the matters of basic justice.¹⁴ A political conception of justice, and justice as fairness is but one of those conceptions, offers a more or less complete normative articulation of such elements. Thus, a political conception offers a way of ranking and interpreting the central elements of a constitutional democracy. It is in this precise sense in which I understand constitutional consensus as being somehow detached from the two-stage rational of justice as fairness. But the two processes –of reaching a constitutional consensus and of elaborating and justifying conceptions of justice- must be seen as taking place in parallel, at least initially. Once a constitution is adopted and therefore constitutional consensus proper is achieved, the process of political justification

¹⁴ A third crucial idea, although not central to this present discussion, is that of public reason.

gathers momentum. As the initial constitutional consensus becomes deeper and wider, the process of political justification becomes more focused since interpreting and furthering on the constitutional essentials requires the contest of particular conceptions of justice. With these remarks in mind, I turn now to examining the features of constitutional consensus.

2.Rawls on Constitutional Consensus.

Thus far we have looked into various interpretative issues surrounding the idea of constitutional consensus. Now we concentrate on two main aspects associated with it. The first aspect is concerned with its content and main features. The second aspect concerns the way in which constitutional consensus can be obtained, or in Rawls's view, has been obtained. Interestingly, he thinks that most, if not all, Western or Western-type democracies have reached this stage in their particular democratic histories. The significance of this is more than apparent and certain considerations will be added below.

A constitutional consensus may be defined as the process whereby a particular polity comes to adopt a constitution that satisfies a certain 'democratic minimum' as a result of its citizens' general allegiance to that constitution. Such a 'minimum' must include: 1)basic procedures for arbitration between contending groups, and 2)basic political rights and liberties.¹⁵ These basic rights and liberties are grouped under the umbrella of the constitutional essentials. I will examine this notion of a democratic minimum in more detail below. Let us now look at the particular features of a constitutional consensus.

¹⁵ Ibid., pp.158-59.

The way in which political liberalism defines constitutional consensus is, perhaps not surprisingly, by contrasting it with overlapping consensus. The differences between the two refer to what Rawls regards as the three main aspects of a consensus. These are its breadth, depth and scope.¹⁶ Under the depth of a consensus, he considers the extent to which the political principles that form the basis of the agreement are, in a sense, self-referential. In other words, whether the principles are taken as such, or whether they are embedded within specific and substantive conceptions of justice. Hence, whereas constitutional consensus is not deep because its terms of reference are the principles themselves, overlapping consensus is deep because the principles at the heart of the consensus are those defined by a particular and substantive conception of justice (such as justice as fairness). The breadth of the consensus is, roughly speaking, its domain. The contrast in this case is between a consensus on the basic rights and liberties that are essentially necessary for the initial functioning of a constitutional democracy, and a consensus on a more extensive list of rights and liberties, that may also include matters of basic justice. Finally, the scope, or focus, of a consensus is the range of liberal conceptions that feature within it. The scope of a constitutional consensus extends to most, if not all, liberal political conceptions of justice. Beyond this, a constitutional consensus does not have a particular focus. An overlapping consensus, by contrast, is focused on either a single liberal political conceptions of justice, or on a subset of all such conceptions. The wider the consensus, the less specific is its focus.

Let us now look at the two main features of a constitutional consensus. Firstly, a constitutional consensus is a consensus, rather than an agreement. This is a key point, for Rawls is persistently careful in distinguishing between an agreement, for

¹⁶ Ibid., pp.164-68.

instance a *modus vivendi*, and a consensus. But what exactly separates a proper Rawlsian consensus from an agreement? There are two kinds of differences. The first difference is that the parties to an agreement are generally motivated by private interests, where interests are here understood in a wide sense. Perhaps it would be more precise to say that the parties are moved by either private concerns, non-public reasons, and/or political reasons of a sectarian kind. The parties to a consensus are, on the other hand, driven by public reasons. In the case of a constitutional consensus, the parties are aware that the principles they are to consent to will inform their society's public charter, and that therefore this consideration must be at the basis of their proposals, discussion and adoption of norms and principles in a constitutional convention. The second difference is that, if the parties to a consensus are moved by public reasons of some kind, then at least a fairly basic form of reasonable pluralism must obtain. The degree of reasonableness of such pluralism will depend on the type of consensus at hand. We could say that, at a minimum, the parties to a constitutional consensus have moved beyond simple pluralism, since they have committed themselves to a certain way of conducting their political affairs and to honour a basic core of rights and liberties.¹⁷

Secondly, a constitutional consensus is a consensus of certain democratic essentials that Rawls calls the 'constitutional essentials'.¹⁸ These are of two kinds: a) institutional essentials¹⁹, that is, principles that define and regulate the structure of government and of the political process, and b) basic rights and liberties. These in turn can be grouped under three headings:

1. political rights: the right to vote and to participate in the political process;

¹⁷ Ibid., pp.163-64.

¹⁸ Ibid., pp.227-30 and JFR, pp.28,32,41,47-49 and 153-4.

¹⁹ I owe this term to Rex Martin.

2. political liberties: liberty of conscience, freedom of thought and speech and freedom of association, and
3. the protection of the rule of law.

Combining the two features, we can say that a constitutional consensus is achieved when:

1. a constitution establishing basic rights and liberties and the necessary democratic procedures has been adopted
2. the content of the constitution, its essentials, has been arrived at through a process of reflection, discussion and negotiation using public reason, and
3. as a result, simple pluralism begins to turn into reasonable pluralism.

The idea of constitutional essentials responds to the notion of what I have called a 'democratic minimum'. The constitutional essentials focus only on one part of the basic structure, on the institutions that regulate the political process and the rights and liberties of citizens. Now, in the aftermath of the publication of the first edition of PL, one of the fundamental worries of early reviewers and commentators was Rawls's apparent sacrifice of the difference principle and, hence, of distributive concerns.²⁰

One of the reasons they gave was that Rawls's list of constitutional essentials looked very much like the first principle. It seemed as if he had primarily concentrated on securing liberty and its priority.²¹ Thus, doubts regarding the relative importance of distributive justice in PL arose because leaving these questions beyond the pale of the

²⁰ See, for instance, Barry in 'The Search for Stability', and Moller Okin, S., 'Book Review', *American Political Science Review*, 87 (1993), p.1010. For a convincing response to these critiques, see Estlund, D., 'The Survival of Egalitarian Justice in John Rawls's *Political Liberalism*', *The Journal of Political Philosophy*, 4 (1996), pp.68-78.

²¹ Perhaps as a result of Hart's well-known criticism in 'Rawls on Liberty and Its Priority', *University of Chicago Law Review*, 40 (1973), pp.534-55.

constitutional essentials appeared to imply that they were somehow secondary with respect to the matters featured in a constitutional consensus. However, the way I interpret this theoretical exclusion differs from the one just mentioned. Rawls's intention, or so I would argue, is not to abandon distributive concerns, but to secure the democratic rendering of liberty. This seems an indispensable step to take before matters regarding distribution can be effectively tackled. I therefore propose to view constitutional consensus as a pivotal moment in a two-fold process that aims to secure: 1) institutional stability, and 2) the democratic rendering of liberty.

To expand a little further. A given democratic polity is normally confronted with three types of fundamental questions: 1) institutional essentials; 2) constitutional essentials proper; and 3) matters of basic justice. Both institutional essentials and constitutional essentials proper must be placed beyond the reach of majority rule. Now a Rawlsian constitutional consensus is signalled by a general, unspecified agreement on basic political rights and liberties. Such an agreement is not deep since it leaves the content and boundaries of these rights not fixed. It is not wide since it does only include a core of basic rights and there might be other rights that could be also counted as basic. And, its scope is limited since it only applies to political procedures, but not to the basic structure of a society. This is why a constitutional consensus constitutes what I have termed a 'democratic minimum'. The driving idea here is that there is a kind of urgency in settling these two types of essentials.²² In the case of the institutional essentials, because frequent changes in the structure of government may become a source of political uncertainty, and may lead to citizen's distrust in the effectiveness of constitutional democracy. In the case of the basic rights and liberties, because they are consubstantial with a free democratic regime.

²² PL, p.227.

There is an important difference between these two types of essentials. The institutional essentials may vary between different constitutional democracies²³; the way in which a particular democratic society takes shape institutionally very much depends on its political traditions and on the specifics of its political history. However, in the case of the constitutional essentials proper, there is not much room for variation. This is because basic rights and liberties can only be stated in one way across democratic societies.²⁴ Furthermore, it is important that basic rights and liberties be formulated in a rather general form, since it is highly undesirable for an essential to become the subject of interpretive controversy. This is why, in Rawls's view, the difference principle is not a constitutional essential, for he concedes that it may be difficult to agree on, for instance, the types of policies that need to be adopted for its adequate achievement. Discussion of such policies is normally conducted by legislative bodies. Hence, the difference principle is not needed to feature in a constitutional consensus as it is 'too demanding'.²⁵ By contrast, a social minimum is just a principle and, as such, may be considered an essential of the second kind. In relation to this, it is far easier to ascertain whether an essential covering basic rights and liberties is realized than it is to do the same with respect to basic justice. Rawls illustrates this by suggesting that matters of basic justice 'are nearly always open to wide differences of reasonable opinion; they rest on complicated inferences and intuitive judgements that require us to assess complex social and economic information about topics poorly understood.'²⁶ Although Rawls does not explicitly mention it, this looks very much like an explanation based on the burdens on

²³ *Ibid.*, p.228.

²⁴ *Ibid.*

²⁵ *Ibid.*, p.229.

²⁶ *Ibid.*

judgement. We can say therefore that matters of basic justice are less immune, if at all, to the burdens of judgement than constitutional essentials.

The preceding discussion has introduced the four requirements of constitutional essentials.²⁷ These are:

1. their political urgency ('It is more urgent to settle the essentials dealing with the basic freedoms.')
2. their normative indispensability for a democratic polity (The basic structure of society 'specifies and secures citizens' equal basic rights and liberties and institutes just political procedures.')
3. their transparency ('It is far easier to tell whether those essentials are realized.')
4. their greater availability to a consensus ('It is much easier to gain agreement about what the basic rights and liberties should be [at least] about the main outlines.')

These three requirements are, in my view, central for the idea of stability of the required kind. In fact, Rawls himself acknowledges that 'so long as there is firm agreement on the constitutional essentials and established political procedures are reasonably regarded as fair, willing political and social cooperation between free and equal citizens can normally be maintained.'²⁸ I think that we are now in a position to effect the change in the argument for the possibility of justice as fairness, and hence to make more explicit the connections between constitutional consensus and stability.

²⁷ Here I follow Rawls's explanation in *ibid.*, pp.229-30.

²⁸ *Ibid.*, p.230.

3.Stability and constitutional consensus.

Having examined Rawls's notion of constitutional consensus, my aim now is to explore what is, and what is not, achieved there in terms of stability. My claim, as I have been suggesting, is that a constitutional consensus affords a significant degree of stability of the required kind where this, as I have also been suggesting, implies institutional stability. Even though we have reasons to interpret constitutional consensus in this way, some difficulties may arise. These difficulties are derived from Rawls's evolutionary account of the path to overlapping consensus. Once these difficulties are overcome, we will be in a better position to show: 1)why a modus vivendi cannot afford the requisite degree of stability, and 2)why constitutional consensus is necessary, but not yet sufficient for stability. First of all, I wish to spell out in more detail why interpreting constitutional consensus in this way signifies an improvement on the framework of political liberalism.

My suggestion throughout this thesis has been that institutional stability may be a more appropriate target than moral stability. The reason is not merely, or not principally, one of futility. Admittedly, pluralism sets the stakes too high for moral stability, but this is not the whole of the story. The main reason why we ought to focus on institutional stability is that the programme of political liberalism requires us to do so. And it requires us to do so because of the project's depiction of the relationship between liberalism and democracy as a mutually supportive one.²⁹ Hence, constitutional consensus needs to be seen as that crucial point at which: 1)liberty becomes democratically institutionalised and 2)democratic institutions embed the principle of basic liberty. Just as basic liberty requires its democratic

²⁹ Amy Gutmann expresses a similar view when she claims that, contrary to those, like Shklar, who considered the relationship between liberalism and democracy as a marriage of convenience, Rawls's (and also Habermas') understanding of this relationship seems closer to that of an intimate marriage, 'Rawls on the Relationship between Liberalism and Democracy', pp.174-80.

rendering so that it does not function in a kind of institutional vacuum, democratic institutions necessitate the underpinning of basic liberty. A constitutional consensus, therefore, signals the endorsement on the part of citizens of the framework of a democratic *Rechtstaat*.

A constitutional consensus is, above all, a theoretical concept. This may seem rather obvious but, at times, constitutional consensus is portrayed as a specific, and particularly important, stage of a polity's actual democratic history (and of US democratic history in particular). Whereas I do not wish to question the historical relevance of constitutional consensus, such identification between a political concept and a particular moment in the tradition of democratic development of one polity (or at the very least of a relatively small number of polities) poses certain problems. For instance, much has been said about judicial review and its usurpation of some of the powers of actual decision-making from the (truly democratic) realm of legislative politics.³⁰ The claim is that, by making the Supreme Court the exemplar of public reason, political liberalism endorses a view of democratic politics that prioritises the role of an unaccountable and unelected judiciary over elected legislative bodies. It is not my intention here to contribute to the debate of whether judicial review is a gain or a loss to democracy. But Rawls's remark that 'in a well-ordered society the two (public reason and the Supreme Court) more or less overlap'³¹ certainly seems problematic. For, if public reason applies specifically to the decisions made by the

³⁰ This literature is vast but some useful sources for our purposes here are, for instance, Waldron, J., 'Judicial Review and the Conditions of Democracy', *The Journal of Political Philosophy*, 6 (1998), pp.335-55, Bellamy, R., *Liberalism and Pluralism*, esp. ch.2, Evans, M., 'Public Reason as Liberal Myth', *Journal of Transatlantic Studies*, 1 (2003), pp.8-25, and Michelman, F.I., 'Rawls on Constitutionalism and Constitutional Law'. The controversy over judicial review is a long-seated one. For works that are critical of the impact of judicial politics on democracy, see Dahl, R., *Democracy and its Critics* (New Haven: Yale University Press, 1989), and Waldron, J., 'A Right-Based Critique of Constitutional Rights', *Oxford Journal of Legal Studies*, 13 (1993). For defenders of judicial review see Freeman, S., 'Constitutional Democracy and the Legitimacy of Judicial Review', *Law and Philosophy*, 9 (1990) and Dworkin, R., *Freedom's Law. The Moral Reading of the American Constitution* (Cambridge, MA: Harvard University Press, 1996).

³¹ PL, p.231.

judges of the Supreme Court, it may seem that such a body is regarded by Rawls as an indispensable element of democracy. However, that this is not his intention becomes clear in the following passage:

Some will say, certainly, that parliamentary supremacy with no bill of rights at all is superior to our dualist regime. It offers firmer support for the values that higher law in the dualist scheme tries to secure. On the other hand, some may think it better that a constitution entrench a list of basic rights, as the German constitution does. It places those rights beyond amendment, even by the people and the German supreme court, and in enforcing those rights can be said to be undemocratic. Entrenchment has that consequence. *Judged by the values of a reasonable political conception of justice, these regimes may be superior to a dualist regime in which these basis questions are settled by the higher law of We the People.*

*Political liberalism as such (...) does not assert or deny any of these claims (...). Our point here is simply that, however these questions are decided, the content of a political conception of justice includes the values of public reason by appeal to which the merits of the three kinds of regime are to be judged.*³²

From this we can conclude that judicial review is merely a contingent instrument of democracy, and one that relates to particular democratic traditions and their histories. What seems far more determinant in delimitating the constitutional essentials is, as we saw above, a requirement of simplicity in their application.³³ To this we may add another important aspect already suggested, namely, that institutional essentials (such as judicial review) can, and do, exhibit a certain kind of morphological variety across regimes. This contrasts with the constitutional essentials proper, which are roughly fixed for all democratic regimes.

This absorption of instrumental aspects into theoretical ones appears all the more clear when we consider Rawls's particular account how a *modus vivendi* might turn into a constitutional consensus. In my view, we need to abstract from this particular brand of evolutionary logic in order to understand more clearly the issue at stake here, namely, whether or not *modus vivendi* can solve stability. There is one

³² Ibid., pp.234-35, emphasis added.

³³ For this interpretation of Rawls's understanding of judicial review (which I largely share), see Michelman, F.I., 'Rawls on Constitutionalism and Constitutional Law', pp.403-06. Michelman argues that Rawls has 'a further motivation for the sorting (of constitutional essentials) beyond the contingencies of judicial review (...), there is (...) a concern for a relative transparency of application of the constitutional essentials, regardless of who is applying them.' p.404.

main problem associated with this evolutionary account. The problem is that it is parasitic upon a particular account of the advent of constitutional democracies. Not all processes leading to the adoption of a democratic constitution have started from a *modus vivendi* of the Rawlsian type.³⁴ This requires that, prior to the constitutional consensus, ‘certain liberal principles of justice are accepted as a mere *modus vivendi* and are incorporated into existing political institutions.’³⁵ This appears to imply that, given that we have both liberal principles and political institutions, the only difference between a *modus vivendi* and a constitutional consensus is that the latter requires the acceptance and affirmation on the part of citizens of the principles themselves as opposed to their mere support for a more or less peaceful state of affairs.

So, why is a constitutional consensus more stable than a *modus vivendi*? Or, in other words, why does a *modus vivendi* fall short of stability of the required kind? Both cover a very basic core of principles, and both are, in a sense, self-referential. This is an important question, particularly in the light of the concerns voiced by pluralists and their claim that anything beyond a *modus vivendi* is too thick to cope with diversity³⁶. The Rawlsian answer is clear: a constitutional consensus is more stable than a *modus vivendi* because it goes some way towards securing a democratic rendering of the basic core of rights and liberties that might have been previously accepted as a result of a *modus vivendi*³⁷.

This can be fleshed out as follows. We start with the basic core of rights and liberties that might have been accepted as a mere *modus vivendi*. Let us call this the basic core of liberal principles. At this stage we do not introduce considerations of

³⁴ It is not my intention to get drawn into a kind of phenomenological account of so-called ‘democratic transitions’; all I am saying is that we should not generalise and abstract from just one case.

³⁵ PL, p.159.

³⁶ For instance, Bellamy, R., *Liberalism and Pluralism*, or Gray, J., *Two Faces of Liberalism* (Cambridge: Polity, 2000).

³⁷ PL, p.161.

any other kind, namely, democratic considerations. The basic core of liberal principles does not have at this stage any particular institutional embodiment. In fact, it could coexist, at least temporarily and provided certain minimum requisites are met, with a variety of other institutional arrangements.³⁸

Once we introduce democratic considerations, a given *modus vivendi* may be stabilized in this way. Citizens, who may be guided by their own comprehensive doctrines³⁹, are led to accept the basic core of liberal political principles and, at the same time, existing institutions are reformed so that they comply with the standards that such principles embody. This will eventually result in the effective regulation of democratic institutions by liberal principles. This is achieved when the basic core of liberal principles fixes the content of a necessary minimum set of basic political rights and liberties (the constitutional essentials). As a result, such essential matters of basic democratic governance are taken out of the political contest. This clearly contrasts with the situation under a *modus vivendi* where these questions are always at the forefront of political conflict.

Secondly, given that the basic core of liberal principles refers strictly to institutional and political facts, liberal principles can be applied following public reason. Compare this with a *modus vivendi* where there is no place for public reason, only for the non-public reasons of different groups and factions. Furthermore, the basic core of liberal principles, by its very nature of being an unspecified core with no attention to detail, is relatively uncontroversial. Most conceptions of justice of liberal persuasion would endorse this minimal liberal agenda as a general interpretation of

³⁸ This is how we need to interpret the passage quoted above since otherwise *modus vivendi* and constitutional consensus would be very similar.

³⁹ The way in which political principles and comprehensive doctrines are connected here is much looser than in the case of overlapping consensus: '...many if not most citizens come to affirm the principles of justice incorporated into their constitution and political practice without seeing any particular connection, one way or the other, between these principles and their other views.' PL, p.160.

the basics of a constitutional democracy. Matters and/or principles requiring more specific elaboration would go beyond a constitutional consensus.

Thirdly, as a result of this process citizens learn to value public reason and they come to realize that although they started within the standpoint of their particular non-public reasons, they have moved beyond that, and have realised that institutionalising a principled common ground is an advantage over previous states of affairs. This also has further advantages like the extension of the cooperative virtues of reasonableness, fairness, a spirit of compromise and the willingness to cooperate on generally accepted terms⁴⁰.

As we have been suggesting, there seems to be a kind of evolutionary logic built into Rawls's account. It starts with an unreflected minimal core of liberal principles with no particular institutional bearing, and then moves to its institutionalisation along democratic lines. The stage of constitutional consensus represents an advance from the point of view of stability. I am not disputing the plausibility of this process here, what I am claiming is that, by relying on this rather narrative account, Rawls has not really provided an answer to our initial question of why a *modus vivendi* is not stable enough.

What I propose is that we abstract from Rawls's account of the story of an evolution and we focus on defining *modus vivendi* and its features more appropriately, and on examining why it cannot solve stability. This is crucial to fend off the objections raised by the pluralists. A first difficulty that we hit is that of defining *modus vivendi*. Here we need to address three questions in particular: what does it entail at the level of principles? What institutional embodiment does it have? And, what type of support does it require?

⁴⁰ Ibid., p.163.

These are crucial questions that require some clarification, not least because of the objections raised by pluralists. Some pluralists, such as John Gray, claim that the stability of a liberal order can only be guaranteed, given pluralism, under a *modus vivendi*⁴¹. The incommensurability of values reflected in the diverse ways of life that are manifested in modern societies requires a temporarily provisional political arrangement that can ensure the peaceful coexistence of those diverse forms of life. Other pluralists, like Bellamy, maintain that anything beyond *modus vivendi* requires a consensus on such fundamental matters as constitutional essentials and pluralism cuts against this possibility⁴². The difficulties of maintaining a general consensus on a constitution are, in Bellamy's view, clearly patent in the controversies surrounding judicial review and in the so-called juridification of U.S. politics. Perhaps pluralists are more optimistic than they should be with regards to a *modus vivendi*'s ability to contain conflict. And the main reason to be sceptical here is that such a kind of political order may be neither liberal nor democratic. This is not explicitly acknowledged by Rawls, although some of what he says can be understood in this vein. Simply put, my point is that a *modus vivendi* cannot solve stability because it cannot secure the democratic rendering of liberty.

A *modus vivendi*, and Gray and Bellamy as well as Rawls would agree on this, is characterized by its institutional provisionality. Gray sees this as a good thing because it encapsulates the foundational spirit of liberalism, which is the spirit of toleration. Bellamy also thinks that having provisional political arrangements is a good thing because this leaves open the possibility of changing them as a result of democratic negotiation and compromise. For Rawls, this kind of institutional provisionality shows a weakness of both liberalism and democracy. The main reason

⁴¹ Gray, J., *Two Faces of Liberalism*.

⁴² See his *Liberalism and Pluralism*.

that Rawls gives against provisionality is that any change in the balance of power within a given *modus vivendi* is bound to affect its stability. Hence stability in a *modus vivendi* is only temporal. This reason, in my view, overlooks a more important difficulty with this notion. In fact, I would go as far as to claim that Rawls's explanation of why *modus vivendi* cannot solve stability is weak, and it is weak because of his reliance on the evolutionary account. Hence, Rawls, as well as Gray and Bellamy, has afforded *modus vivendi* a liberal pedigree that it may not necessarily have, for let us not forget, he starts his account of how a constitutional consensus may come about by stating: 'Suppose that at a certain time (...) *certain liberal principles of justice are accepted as a mere modus vivendi, and are incorporated into existing political institutions.*'⁴³ This account seems to trace the normative connection between liberalism and democracy back to their historical connection and, therefore, it is incomplete.

From the Rawlsian statement just given, we can extract two differences between *modus vivendi* and constitutional consensus. As I have said, in both (*modus vivendi* and constitutional consensus) liberal principles are affirmed. However, whereas in a constitutional consensus the endorsement of liberal principles is based on public reasons of some kind, in a *modus vivendi*, citizens are guided by their own interests with their only goal being that of reaching a kind of compromise that may contain civil strife. The second difference is even more crucial. A *modus vivendi* is silent with respect of the particular character that its institutions might exhibit. There are no institutional requirements, democratic or otherwise, beyond that of providing peaceful coexistence. This is a key aspect of *modus vivendi* that Gray and Bellamy ignore, and that Rawls himself does not adequately stress.

⁴³ PL, p.159, emphasis added.

Hence, at the level of principles, a *modus vivendi*'s initial liberal character might be reversed by a change in the balance of forces that sustain it and by its institutional provisionality. This institutional provisionality questions its assumed, but not theoretically established, democratic credentials and because of these two features, *modus vivendi* lacks a general commitment on the part of citizens to either a set of liberal principles or to a particular procedure for making political decisions. To be sure, the main feature of a *modus vivendi* is not just the lack of public reasons, as Rawls suggests, but the inexistence of a more or less developed conception of a public political order, as one where both principles and institutions work in tandem to secure the freedom and equality of its citizens.

Now, it may be argued that Rawls's understanding of *modus vivendi* differs in important ways from that of Bellamy, and that the latter gives this notion a more substantially democratic flavour. After all, Bellamy is strongly committed to democratic decision-making even at the level of constitutional essentials. But this is precisely why I proceeded to define the features of a *modus vivendi*. There might be, however, a kind of nominalistic obstacle in looking at what *modus vivendi* entails. In other words, Rawls and Bellamy may have used the same nominator to imply different concepts. There is one important difference between the two theorists in this respect and, arguably, a key one. For Bellamy, a *modus vivendi* requires an institutional commitment to democracy –the commitment to majoritarianism. However, if this is the only institutional condition, his *modus vivendi* can still be subjected to the same charge as the Rawlsian one, for how exactly can majoritarian rule secure the stability of liberal political principles? He seems to have forgotten the potential dangers of pure majoritarianism.

So far I have argued that since our concern is stability, understood as the institutional stability of both liberal principles and democratic procedures, we need to move beyond a *modus vivendi*. Here I have only fleshed out in some detail the implications of *modus vivendi* politics so that Rawls's standpoint can be more adequately defended against the pluralists' views. However, this also constitutes an argument for constitutional consensus as a necessary mechanism for institutional stability. Beyond what we said in the second section of this chapter with respect to the features of constitutional consensus, we need to emphasize some important aspects:

1. It does seem clear that pluralists have, at least implicitly, ignored some of the important difficulties that we have suggested here. They appear to have overestimated the capacity of *modus vivendi* to deal with the question of pluralism. As I have argued, an adequate treatment of the fact of pluralism requires a consensus on principles and institutions and, in particular, on liberal principles and democratic institutions on the basis of their mutual support. But it also requires citizens' allegiance to both principles and institutions, and it is this allegiance to a constitution that fixes certain matters and leaves them off the political agenda (the constitutional essentials) which pluralists like Bellamy view as problematic because it shows a disregard for the very essence of politics, which is not the constitution, but the realm of discussion, negotiation and compromise. It is true that Rawls's remarks on the necessity to fix 'once and for all' some constitutional essentials appear somehow worrying.⁴⁴ But we might have some good reasons for relativising this reference to a permanent fixing of constitutional essentials. Again, this is another instance in which the detail does not let us fully grasp the crucial point and, again, this is partly due to Rawls's reliance on the U.S. experience whose Constitution works not only as a

⁴⁴ PL, p.232

political charter, but as something that, in the absence of a long history and mythology, cements the nation and binds people together politically. But I say that this is only 'partly' due to the generalisation from the one case because the term 'essentials' itself may be ambiguous. One possible way to understand it is by reference to something (in this case principles) that is permanent, fixed and never changes. We can nevertheless understand the 'essentials' as something that is required for the effective functioning of a liberal democratic order. In this sense, the 'essentials' as such may change, but there will always be a core of essentials without which a liberal democratic polity would not be able to function adequately.

2.As an alternative to 'fixing', I have suggested the notion of a 'democratic minimum'. We arrive at this notion by asking: 'what kind of institutional arrangements may be sanctioned so that free and equal people who differ importantly about matters of value, including matters of political value, may peacefully live together? An answer to this question requires a specification of bare liberal principles such as liberty of conscience, freedom of thought and association, freedom of speech, freedom of movement, the protection of the rule of law and some political rights such as the right to vote and to participate in politics. These are more or less the essentials of all liberal democracies, essentials that are universal in their application, that is, they apply equally to all adult citizens in a constitutional democracy. It also requires the adoption of institutions or, perhaps more likely, the reformation of pre-existing ones so that they meet a certain (minimum) democratic standard. Such institutions may be regarded as the set of institutional essentials.

3.A constitutional consensus, therefore, appears indispensable for institutional stability. But is it also sufficient? It might be said here that if my worry is institutional stability, and a constitutional consensus can provide it, why go any further? There are

at least three reasons why we should go further. These reasons derive from our concern with 1)stability, 2)democracy and 3)justice. Our concern with stability demands that we do not stop at institutional stability. The long-term viability of democratic institutions also depends on the attainment of political stability. A polity attains political stability if its institutions, in their day-to-day running, do not depart from the values that ground those very institutions. A constitutional consensus affords the initial institutional stability, but this has to be sustained in the regular functioning of democratic politics. Hence, institutional stability might falter if institutions are put under pressure by legislative politics. But, how can institutions be put under pressure? The explanation for this takes us to the second of our concerns, that of democracy. Since constitutional consensus stops with that 'democratic minimum' of constitutional and institutional essentials, legislative politics takes over the task of dealing with matters of basic justice and with furthering the essentials (of broadening and deepening the consensus). Constitutional essentials are bare principles and they need to be fleshed out in more detail, and perhaps adjusted so that there is no conflict amongst the essentials. Take, for instance, freedom of speech. As an essential, freedom of speech is a bare principle. However, we saw in chapter 2 how, even this bare principle may be highly controversial. The broadening and deepening of the consensus may require certain revisions and adjustments, but these have to be effected in a way that does not compromise the liberal and democratic credentials of those institutions. The difficulty is that, although there is a prima facie case for the overridingness of the constitutional essentials, there may be some instances in which they have to be limited. This process of adjustment, however, will need to fulfil certain criteria so that the democratic rendering of liberty is preserved. I will elaborate on this point in my next chapter.

The preceding discussion takes us to our final concern, which is a concern with justice. In response to the question about the sufficiency of constitutional consensus, we need to say that constitutional consensus remains silent on issues of justice. So far all we have guaranteed is an initial democratic stabilization of liberty. There are no conceptions or principles of justice doing any work at the level of constitutional consensus, at least initially, beyond bare liberal principles themselves. There are two important aspects here. One is that, as political philosophers, we are legitimised to put forward our various conceptions of justice, even those that are pitched at a comprehensive level. This is what Rawls does, although as I have been arguing throughout this thesis, his account of how his conception of justice might be realized in practice suffers from serious deficiencies. Such deficiencies nonetheless do not invalidate the richness of his overall project. The second crucial aspect is that legislative politics requires a focal point of view so that decisions may be made without deviating from the democratic minimum. A conception of justice provides that focal point for the widening and deepening of constitutional consensus. Without such focal point, a given constitutional consensus might revert back to institutional instability. Hence, the project beyond constitutional consensus is that of articulating and justifying conceptions of justice that can serve as a basis of democratic decision-making. And, given pluralism, I have argued that we should strictly present political justifications for our conceptions of justice.

CHAPTER 6

CONSENSUS AND JUSTIFICATION

This chapter deals with the question of how should justice as fairness be justified so that it may be regarded as a realistic utopia. Our previous chapter concluded that constitutional consensus is a necessary, but not yet sufficient, condition of stability of the required kind, where this was understood as institutional stability. This chapter aims to show that free-standing justification is also necessary for institutional stability, and furthermore, that it is both necessary and sufficient for the right kind of justification, namely, political justification. The chapter is divided into five sections. Firstly, I expand on what was said on my previous chapter about the need to go beyond constitutional consensus. In the second section, I look at Rawls's approach to justification and provide a brief assessment. My next section is devoted to the key notion of public reason and its role in the process of democratic justification. I then turn to how justice as fairness might be politically justified following public reason. Finally, I conclude with some reflections on legitimacy and stability.

1. The limits of constitutional consensus.

I concluded my previous chapter by suggesting why we should not rest content with constitutional consensus. I argued that there are three reasons -that relate to stability, democracy and justice- that push us beyond constitutional consensus. Before I elaborate on these, I wish to bring forward one objection to the view defended here (which is also Rawls's position), namely, that we should consent on a constitution before agreeing on justice. This objection, raised by Waldron, states that citizens

cannot agree on issues of constitutional design without prior agreement on the 'telos' of such arrangements.¹ This would seem a sound comment particularly given that Waldron seems to be concerned with the narrowness of Rawls's conception of pluralism. The idea is that disagreements, particularly about justice, should play a full part in the design of a society's constitution. This objection, however, is pitched at the wrong level, for disagreements about justice are sufficiently 'thick', and they can compromise a society's consent on a democratic minimum. Another way of saying this is that justice requires democracy. There may be two types of disagreement about justice: fundamental disagreement and disagreements about matters of detail. A fundamental disagreement would be one between, for instance, a Marxist-Leninist and a liberal. The question here is whether a Marxist-Leninist should present her view of justice in support of a particular type of constitutional arrangement. A Rawlsian take on this is that constitutional arrangements must be justified on the basis of reasons that all can accept as reasonable and rational. And, given what we now know about Marxist-Leninists, it is highly unlikely that their version of a constitution would pass this test of acceptability. The reason for this is not because there may be few Marxist-Leninist in society, but because the grounds for justifying their specific institutions do not meet the test of reciprocal justification². With respect to the second type of disagreement -disagreement about detail-, the Rawlsian position here is that such matters -like basic justice- are not matters to be considered within a constitutional consensus. Hence, this less fundamental type of disagreement does not have much to say about constitutional design either.

We also said that one of the reasons why we should not stop at constitutional consensus is because it might be put under strain by adding to the constitution. So,

¹ Waldron, J., 'Disagreements about Justice', *Pacific Philosophical Quarterly*, 75 (1994), pp.372-87, p.378.

² Reciprocal justification derives from the principle of legitimacy.

under what circumstances can a society that has attained (institutional) stability revert back to instability? As far as citizens are concerned, there may be two reasons for this: i) the endorsement of a constitutional consensus may be shallow, and a fuller endorsement may require a certain ethos of citizenship, and ii) citizens may not be able to distinguish properly between political and non-political values. For instance, some citizens might have approved a constitutional consensus without having reflected on their consequences, particularly on the demands that it may place on them. Note that this differs from the case where a consensus on the constitution is endorsed as a result of *modus vivendi* reasons (i.e., from a self-interested perspective), or from the case where a consensus is accepted for the right reasons and then citizens fail to live up to it (for whatever reasons). The situation that I am referring to here presents some similarities with the one described by Klosko,³ where citizens support constitutional principles and the political system in the abstract, but are not able to see fully how their own political behaviour might be constrained. Because of this, political stability may require an ethos of citizenship. Similarly, in a constitutional consensus, citizens might not have effected a thorough separation between political and non-political values. Again, an ethos of citizenship may be necessary here. Interestingly, it is precisely this kind of problem that leads Rawls towards overlapping consensus. He thinks that this kind of shallow and unreflected support may pose a threat to stability. In his view, an ethos of citizenship is not in itself sufficient to secure stability of the required kind. For Rawls, only full moral stability can guarantee political stability. It is not clear, however, why the ethos of citizenship requires further grounding (in comprehensive doctrines). The perception here seems to be that citizens with a free-standing conception of their ethos as citizens would somehow be less committed to

³ See Chapter 2.

honour that ethos than citizens with an embedded conception. Citizens with a free-standing view of their political status may lack the appropriate kind of motivation to act from this ethos. I am not sure why Rawls thinks that citizens with a free-standing view of their own citizenship might falter. But the problem is this: citizen A might have grounded her political values on a particular comprehensive doctrine, for instance, Catholicism. But, what happens if, after the grounding has taken place, A ceases to endorse the same comprehensive doctrine? It would seem that if the grounding doctrine goes, the ethos of citizenship goes with it. This is, however, contrary to Rawls's conception of the person, and particularly to his view regarding the capacity of citizens to revise their particular ends.⁴ Citizens' political identities remain unaffected by alterations within their non-public identities. As Rawls puts it 'on the road to Damascus Saul of Tarsus becomes Paul the Apostle. Yet such a conversion implies no change in our public or institutional identity...'⁵ The implications of this view seem far wider than Rawls is prepared to admit. For, if the public identity remains the same, the ethos of citizenship will also remain unaffected. The ethos of citizenship is also free-standing.

From the point of view of the political process itself, constitutional consensus only offers a necessary (but minimal) framework from which citizens are able to make binding political decisions. As a minimal framework, constitutional consensus needs expanding. In terms of stability, we may say that to the initial institutional stability achieved there, we need to add political stability so that the framework of institutions to which citizens have consented becomes fully stabilized. In other words, political stability is a requirement (a necessary condition, if you like) for full institutional stability. And, we said that political stability is achieved when a society's political

⁴ PL, pp.30-32.

⁵ Ibid., p.31.

institutions, in accordance with their normal functioning, do not depart from the values that ground those institutions. A constitutional consensus is therefore too 'thin' to oversee the process of further democratic functioning that it initially sanctioned. Such democratic functioning could be put under strain by, for instance, laws that depart from the values of freedom and equality, and also, by citizens and, particularly, by public officials advocating political views contrary to those values.⁶ An example of a political view held by some citizens and public officials that could substantially weaken institutional stability, if it was held by not a negligible number of citizens, is the advocacy by the British National Party of the compulsory repatriation of non-white immigrants and their descendants.

The interesting point here is that different constitutional democracies interpret the constitutional essentials differently. For instance, in the British case, the right of free speech usually prevails, there is a legal tolerance with regards to the expression of the kind of view mentioned above. Other constitutional democracies may effect a different interpretation of the constitutional essentials. Denmark is a useful example here. The recent case of the historian David Irving, who was convicted of Holocaust denial, highlights how the right to free speech may be overridden for the sake of freedom itself. The conclusion from all these examples is the same. There may be many routes to political stability, some may require more institutional intervention than others, but all of them necessarily go beyond constitutional consensus.

Beyond considerations of institutional and political stability, our concerns with justice also militate against the possibility of stopping at a mere consensus on constitutional arrangements. This is an argument, not so much from the requirements of democratic politics, but from political philosophy itself. Our capacity as

⁶ To this we may add that institutional stability may be compromised by judicial decisions that depart from those values.

philosophers to abstract from the ordinary conditions of every day living may result in the elaboration of fairly specific conceptions of justice. Some of these conceptions will be pitched at a comprehensive, or partially comprehensive, level. There is nothing disturbing in this claim. We saw in chapter 3 the shortcomings of justice as fairness with regards to the basic claims of justice advanced by feminists. Perhaps justice as fairness is not, all things considered, able to make sense fully of the extent to which our political and social relationships are determined by the gendered structures in which we live. Whether this is the case or not –in other words, whether justice as fairness can really deliver justice for women- need not concern us here. The point is, rather, that our theoretical efforts to establish how justice should be delivered should be not confused with the similar, but not identical, question of how we should, as free and equal citizens, peacefully live together. Living together peacefully, even as free and equal, is not the same as living together justly. Of course, it might be said that the pursuit of justice may compromise the extent of peaceful coexistence (this seems to have been Rawls's worry). But, all we can say here is that it need not do so. It is precisely this juncture between justice and peaceful coexistence (of free and equal citizens) that Rawls wishes to explore and extend as far as possible. This is a legitimate move but it implies that that foundational moment of peaceful coexistence of free and equal –constitutional consensus- be surpassed. Hence, institutional and political stability are prior requirements of our further quest for justice. And, because the pursuit of justice necessitates political stability so that political decisions do not compromise the stability of the (constitutional) institutional order, at a theoretical level, constructing justice for this kind of political order requires that arguments for justice proceed from values and arguments that may be politically justified. I turn now to Rawls's approach to justification.

2. Rawls on Justification.

As we have seen, in the 1993 edition of PL, as well as in the relevant papers prior to its publication, overlapping consensus is linked, not to the question of justification, but to the question of stability. An overlapping consensus is mainly designed to settle his long-standing concerns with stability, and it is theorized as a necessary condition for the stability of justice. It is not until the 1995 'Reply to Habermas' that the two ideas of overlapping consensus and justification appear closely connected. To the main question directed by Habermas at the idea of an overlapping consensus -whether it substantively adds to the justification of the favoured conception of justice already justified as free-standing-, Rawls's answer is to break down the process of justification of justice as fairness into three stages: i) *pro tanto* justification, ii) full justification by individual citizens; and iii) public justification by political society.

Pro tanto justification refers to the justification of justice as fairness from political values alone. This corresponds roughly with the free-standing stage. The second and third stages both relate to a different type of justification of justice, where justice as fairness is also sanctioned from the different non-political, and non-public, values of the relevant polity *for the right reasons*.

Before I go on to examine *pro-tanto* or free-standing justification, Let me make some observation on the second and third stages of justification. Rawls claims that 'since political justification is *pro tanto*, it may be overridden by citizens' comprehensive doctrines once all values are tallied up.'⁷ This poses a clear problem, as we have seen. The problem is that, in the absence of an overlapping consensus, the whole justificatory strategy fails. This is why we need to provide a substantially closer link between stability and political justification. The second and third stages of the

⁷ PL, p.386.

justificatory process both relate to, what we have called, a moral justification of justice, where justice as fairness is also sanctioned from the different non-political, and non-public, values of the relevant polity *for the right reasons*. They seem to have been thought of in a kind of sequence. Hence, in the first instance, it is the task of each citizen individually to work out how a particular non-public viewpoint relate to the political conception. Here citizens make particular connections between the political conception and their own non-political values from the point of view of their own comprehensive doctrines. The political conception does not dictate how this is done. Rawls's premise here is that a political conception of justice, although freestanding, can 'be embedded in various ways –or mapped, or inserted as a module- into the different doctrines citizens affirm.'⁸ Once this task is successfully accomplished, we move further from the non-public point of view and look for a general public justification of justice. Here we deviate from the comprehensive doctrines that have supplied the grounds for endorsing the political conception, focusing instead on the mere fact of the existence of a consensus on it, albeit for the right reasons. This third stage signals a *public justification by political society*, and Rawls here connects our familiar triad of ideas: overlapping consensus, stability for the right reasons, and legitimacy. Hence a political conception may become fully publicly justified 'when all the reasonable members of political society carry out a justification of the shared political conception by embedding it in their several reasonable comprehensive views.'⁹ By contrast with the full justification, the particular content of comprehensive doctrines does not influence the process. The political conception becomes publicly justified when it is 'the common ground and all reasonable citizens taken collectively (...) are held in general and wide reflective equilibrium in affirming the political

⁸ Ibid., p.387.

⁹ Ibid.

conception on the basis of their several reasonable comprehensive doctrines.¹⁰ What we need to say here is that, implicit in this three-stage view of the justificatory process, there is a presupposition that an overlapping consensus will obtain. But, if overlapping consensus is not a fact, then neither public nor full justification can be guaranteed. We could grant, however, that it may be useful to examine whether and how the political values of justice as fairness can relate to citizens' overall domain of values, but this would entail a different kind of exercise. Such kind of enterprise would be one of critical justification. And, critical justification is moral, rather than political.¹¹ Hence, neither public nor full justification can be considered types of political justification, and neither of them is required for stability.

In actual fact, *pro tanto* justification does a lot of the work as far as justification goes. This can be clearly seen in the arguments for the basic liberties and their priority.¹² Rawls mounts his defence of his two principles of justice on the grounds that they provide a more adequate understanding of the values of freedom and equality in a democratic society than the principles that belong to traditional doctrines such as utilitarianism and perfectionism.¹³ The reason why the two principles provide this superior understanding is because they connect more closely (than utilitarianism, say) with the idea of fair terms of cooperation for the mutual advantage of citizens conceived as reasonable and rational. Basic liberties are regarded as 'essential social conditions for the adequate development and full exercise of the two powers of moral personality over a complete life.'¹⁴ To explain why the basic liberties are central to

¹⁰ Ibid., p.388.

¹¹ This suggestion will be examined in my final chapter.

¹² PL, Lecture VIII. This is an extended version of the Tanner Lecture of the same title given in April 1981.

¹³ Ibid., p.292.

¹⁴ Ibid., p.293.

justice, Rawls provides his account of primary goods.¹⁵ Primary goods are envisaged as ‘citizens’ needs’. They are destined to ensure that citizens can adequately develop their two moral powers, their capacity for a sense of justice, and their ability to pursue their conceptions of the good.¹⁶ Citizens’ two moral powers provide the means of effecting a distinction between the basic liberties, and other non-basic liberties. To be certain, Rawls does not provide an exhaustive list of basic liberties, but this need not concern us here. I am only trying to show that Rawls deploys strictly political arguments for the justification of the basic liberties.¹⁷

The justification for the basic liberties –a process that is roughly completed at the constitutional consensus level- proceeds from its connection with three specific grounds. The basic liberties are requirements of 1)people’s ability to pursue their conceptions of the good (second moral power), 2)their ability to form and apply a sense of justice, and 3)politically securing other rights and liberties.

Because of an eventual conflict among the basic liberties, the institutionalisation of such liberties must be altered so that they form a coherent system, whereby one liberty can only be limited for the sake of another liberty, or set of liberties. Such a scheme of basic liberties needs to be specified, and then adjusted at later stages. The important point is that arguments need to take into account the two moral powers of persons and the primary goods. Rawls also considers a criterion of significance whereby a liberty is more or less significant ‘depending on whether it is more or less essentially involved in, or is a more or less necessary institutional means to protect, the full and informed and effective exercise of the moral powers.’¹⁸ Claims

¹⁵ PL, pp.178-90.

¹⁶ PL, p.187.

¹⁷ For a more extensive list that expands on Rawls’s framework, see Nickel, J.W., ‘Rethinking Rawls’s Theory of Liberty and Rights’, *Chicago-Kent Law Review*, 69 (1994), pp.763-86.

¹⁸ *Ibid.*, p.336.

to particular freedoms must be adjudicated according to their significance so defined.

This applies to the constitutional stage as well as to legislative procedures.

Take, for instance, liberty of conscience and liberty of association. These are protected by appropriate constitutional devices. However, they must be more specifically articulated at the legislative stage. The grounds of these two liberties are provided by the second moral power, in particular by the revisability argument and the internalisation argument.¹⁹ According to the second moral power, persons are self-authenticating sources of valid claims, they are able to make claims on institutions on the basis of their conceptions of the good. But, as citizens, persons are conceived independently of their various ends and attachments. They are also able to revise, alter or change their conceptions of the good. This feature of the first moral power grounds liberty of conscience. This entails no loss to the person's public identity.

The first moral power grounds the basic liberty of free political speech. And, it is not difficult to see why. An adequate materialization of a citizen's sense of justice requires, for instance, appropriate access to informative resources. Freedom of speech entails the protection of advocacy of revolutionary and seditious doctrines, and also, that there be no pre-defined limits to the freedom of the press.²⁰

The important point here is that, since these liberties are elaborated in the abstract, they need to be adjusted in the legislative stage and, in doing so, the same arguments used in their justification, should be used in their relative adjustments and applications. In this process, the political conception of the person and the two moral powers, together with the account of primary goods need to have a fundamental place.

I have provided this outline with the aim of showing that *pro tanto* or political justification performs an important role within Rawls's strategy of justification. The

¹⁹ Ibid., pp.29-35.

²⁰ To this Rawls adds that in a system of basic liberties, there is no such thing as the crime of seditious libel. Ibid., p.342.

justificatory programme contains exclusively political values and ideas. Moreover, this same programme is to be carried through beyond the initial justification of the basic liberties (at the constitutional stage) to their further elaboration, adjustment and application. My next section is devoted to Rawls's important idea of public reason.

3.The Idea of Public Reason.

The idea of political liberalism is central to the project of political liberalism. Since here we are concerned with the question about what type of political argument is appropriate to use for the justification of justice in a democratic context, we are clearly led towards this idea which is 'part of the idea of democracy itself'²¹. Theorizing about public reason comes as a result of asking a crucial question: what kinds of reasons should citizens give one another when fundamental political questions are at stake.²²

Public reason may be defined as the reason of political society, the way political society has of 'formulating its plans, of putting its ends in an order of priority and of making its decisions accordingly'.²³ Thus, public reason refers to the way in which political society tries to make binding decisions on basic political questions in accordance with certain notions and principles. Public reason is said to be public in a threefold sense: it is the reason of the public (of citizens), its subject is the good of the public (constitutional essentials and matters of basic justice) and its nature and content is public (given by the ideals and principles contained in the political conception of justice).

²¹ IPRR, p.573.

²² Ibid., p.574.

²³ Ibid., p.212. In this discussion, I generally draw on Rawls's two most complete elaborations of this idea: Ibid., Lecture VI and IPRR.

More than an idea, public reason is an *ideal* connected to a democratic conception of citizenship. In this respect, public reason 'presents how things might be, taking people as a just and well-ordered society would encourage them to be. It describes what is possible and can be, yet may never be...' ²⁴ As public reason is part of a political conception of justice that is, broadly speaking, liberal, it differs from other modes of reasoning that are non-public, but that are nonetheless not private. Rawls refers to these kinds of reasons as non-public or, sometimes, social. They are strictly non-political reasons, although they are public in a certain sense. Amongst these he mentions explicitly the reasons of Churches and Universities. These, among others, are regarded as part of the background culture of civil society. Public reason is also an ideal in the sense that, when citizens give one another public reasons, they fulfil their (political) duty of civility that results in civic comity. In the absence of public reason, the great good of the political values may be replaced by the great evil of civil discord. Civic comity is therefore required to avoid civic discord.

As a democratic idea characteristic of a democratic people, public reason not only applies to the political, understood as the institutional. It also, and more specifically, applies to citizens when they engage in political advocacy, and when they exercise the right to vote. However, as an ideal, it might be too demanding to expect citizens to follow public reason in the settlement of all, or most, political matters. Thus, the question arises, when does it appear necessary that public reason be invoked? The answer given by Rawls is that it is necessary that citizens honour public reason when issues regarding constitutional essentials and matters of basic justice are involved. It may not be possible to extend public reason to cover other political

²⁴ PL, p.213.

questions. However, citizens fulfil the duty of civility when they make government officials accountable from the point of view of public reason.

As an ideal, public reason needs to be specified and adapted to varying circumstances. Hence, there is some interpretive scope for public reason. In a constitutional democracy with judicial review, the Supreme Court is given this role of interpretation. In judicial review, judges interpret the constitution. They do so by reasoning about the constitution and also from the constitution, offering, therefore, constitutional reasons. Hence, Rawls regards this judicial reasoning from judicial review, as an exemplar of public reason.

The requirement to follow public reason arises out of an idea of democratic citizenship.²⁵ Rawls views the citizens of a democracy as endowed with a kind of civic sense (civic friendship). It is the responsibility of citizens to conduct their political affairs in a way that takes into account and incorporates the social perspective. Citizens' actions in the public forum must somehow be informed by ideas of what is good for all. This is what is entailed by the duty of civility. The duty of civility is an 'intrinsically moral duty'²⁶ that requires that citizens use public reason as a kind of common language for public political advocacy, that they acquire certain dispositions such as the willingness to listen to fellow citizens, and that they recognize the necessity to make reasonable accommodation, should it arise. The question is thus the following: 'by what ideals and principles (...) are citizens to share equally in ultimate political power to exercise that power so that each can reasonably justify his or her political decisions to everyone?'²⁷ An answer to this questions requires that citizens deploy a criterion of reciprocal justification: those proposing fair terms of cooperation must think that it is reasonable for others to accept them, as free and equal. And, this

²⁵ IPRR, pp.577-79.

²⁶ Ibid., p.577.

²⁷ Ibid., p.578.

justificatory reciprocity is, in turn, required by the liberal principle of legitimacy. This principle as applied to public reason states that: 'our exercise of political power is proper only when we sincerely believe that the reasons we would offer for our political action (...) are sufficient, and we also reasonably think that other citizens might also reasonably accept those reasons.'²⁸ The criterion of reciprocal justification applies to the constitution, and also, to the laws enacted from that constitution. Democratic citizenship and legitimacy sanction that only constitutions that satisfy the principle of justificatory reciprocity are reasonable.

When citizens use public reason, they must use principles and guidelines that satisfy the criterion of reciprocal justification. Beyond this limit, the content of public reason is provided by a family of liberal political conceptions of justice. Justice as fairness is only one amongst this family of conceptions. A public political culture may contain competing conceptions of public reason. Even if one, or a few, conceptions were to dominate, there would always be a set of permissible conceptions of public reason; this is important for the possibility of political dissent.²⁹ Public reason is therefore not fixed 'once and for all'.

Public reasoning is always reasoning from within a political conception of justice. It uses political values as distinct from other types of public values and as they apply to the political. Political liberalism only considers those political conceptions that have a place in a constitutional democracy. Furthermore, each of the different conceptions of public reason needs to be complete. Each conception elaborates and presents a self-contained view of public reason and indicates how principles, ideas and guidelines of inquiry can be ordered so that they can provide answers to all, or almost all, political questions. This is a crucial aspect of public reason, for 'unless a political

²⁸ Ibid.

²⁹ Ibid., p.583.

conception is complete, it is not an adequate framework of thought in the light of which the discussion of fundamental political questions can be carried out.’³⁰

So far, this account of public reason has implicitly assumed that it is possible for citizens to separate their public from their private identity. It appears a highly demanding conception of how citizens should conduct their political arguments in the public forum. But public reason does not exclude comprehensive doctrines from public political discussion. Arguments from comprehensive doctrines may be given, provided that, at some point in the future, strictly political reasons are supplied in support of those same arguments already introduced by comprehensive doctrines. This is known as the *proviso*.³¹ However, even though comprehensive doctrines, religious and secular, may have a place in the public political culture, this does not alter the way in which public reason should be justified, i.e. from the family of liberal political conceptions of justice. The *proviso* is required because public justification is not merely reasoning from some premises to their conclusion, but essentially ‘argument addressed to others’.

A final crucial point is that public reason does not necessarily issue in a specific agreement. But this does not entail a failure on its part. When citizens follow public reason, they ‘profit from debate and argument, and when their arguments follow public reason, they instruct society’s political culture and deepen their understanding of one another even when agreement cannot be reached.’³²

³⁰ Ibid., p.585.

³¹ Ibid., p.591.

³² Ibid., p.607. The idea of public reason has given rise to a number of objections. See, for instance, Horton, J., ‘Rawls, Public Reason and the Limits of Liberal Justification’, *Contemporary Political Theory*, 2 (2003), pp.5-23, Quong, J., ‘The Scope of Public Reason’, *Political Studies*, 52 (2004), pp.233-50, and Reidy, D.A., ‘Rawls’s Wide View of Public Reason: Not Wide Enough’, *Res Publica*, 6 (2000), pp.49-72. J. Quong argues that there are important reasons why public reason should be extended beyond the constitutional essentials and matters of basic justice. J.Horton suggests that the principle of legitimacy is too demanding and that it is questionable whether public reason can deal with the level of political disagreement within constitutional democracies. D. Reidy doubts whether public reason can lead to a determinate resolution of fundamental political questions within public reason

4. Public Reason and Political Justification.

Here I am concerned with the question of whether it is possible to effect a political justification of justice as fairness from the preceding account of public reason. A crucial question is the following: under what circumstances would citizens, guided by public reason, consent to adopt justice as fairness as their regulative conception of justice?

When looking at Rawls's theorization of public reason, one important point is to discern which parts belong to the concept of public reason itself, and which parts belong to the conception of public reason of justice as fairness. Without any specific guidance in this respect, we presume that Rawls has simply laid out the conception of public reason of the family of liberal political conceptions or, in other words, the conception of public reason of liberalism in general. We need to assume, therefore, that all the features that we have mentioned, apply equally across each and all of the liberal conceptions of public reason.

There are two ways of looking at this question. One is to examine whether, using the procedures of public reason and on the basis of political values alone, we may be able to get from certain concerns about justice to justice as fairness. The second option is to see whether citizens themselves, when thinking about justice, would all, following public reason, arrive at a result on justice as fairness. The first option assumes that public reason has 'a life of its own', and that it is an ideal, to a certain extent, independent from those who exercise it. According to this view, public reason is a rather mechanical matter that citizens need to 'follow'.³³ However, I think that, on the basis of what was said above about public reason, the second option, of

alone. It is not my purpose to discuss these views here. I will consider the issue of legitimacy in the final part of this chapter. Some of the problems mentioned by these critics will be implicit in my discussion of how justice as fairness may be justified from public reason.

³³ IPRR, p.580.

letting citizens present arguments that fulfil the criteria of public reason, would appear to reflect more accurately Rawls's position.

There is a second difficulty involved in trying to effect a political justification of justice as fairness from public reason. This difficulty concerns whether it should be shown that justice as fairness 1) would be adopted by citizens, 2) can be adopted by citizens, or 3) may be adopted by citizens. The first option needs to be discarded almost immediately, for it is difficult to show how, given reasonable disagreement, citizens would necessarily consent to justice as fairness. The second option should also be discarded because it is rather weak. To claim that justice as fairness can be adopted is similar to claiming that 'I can run 200 metres in 40 seconds'. It does not justify anything. Hence, we are left with the third option. We need to be concerned with how citizens, using public reason, may adopt justice as fairness.

The third difficulty connects with the scope of justice as fairness. It is not clear what exactly is the focus of the possible consent. Is it necessary to show that citizens may adopt, for instance, the political conception of the person, the burdens of judgement and the idea of reasonableness? Or, is it sufficient to show that citizens may adopt the two principles? Also, there are practical matters of space here, and to show how justice as fairness as a whole may be endorsed could prove quite a daunting exercise. My aim is therefore to show how citizens may adopt the difference principle. This is because the difference principle is one of the features that clearly distinguishes justice as fairness from the family of liberal conceptions. Furthermore, the difference principle belongs to the matters of basic justice that should be settled on the basis of public reason.

We start by considering citizens 'as they are'. We may think of citizens as holding a variety of ideas and values that generally relate to economic justice. Some

citizens may have fully or partially comprehensive views, others may have pluralist views, yet other may hold free-standing values and principles. Hence, we need to show how citizens may proceed from these different understandings of economic justice to the difference principle. The aim here is to establish justice as fairness as the main candidate for the interpretation and articulation of distributive economic policies at the legislative stage.

A possible rationale for citizens' selection and further justification of the difference principle could be the following:

a) firstly, the society to which the difference principle is to apply would need to offer both short and long-term evaluation of its economic prospects. It may be difficult, given the burdens of judgment, to come up with a diagnosis that most citizens would accept, but arguments for any principle of redistribution should not ignore the general economic conditions of society, and also of other relevant social conditions.

b) secondly, citizens would need to be presented with the different candidates; here candidates might range from distributive principles proper such as the difference principle, or the idea of an unconditional basic income³⁴, to candidates that offer a more 'comprehensive' elaboration of the requirements of economic justice, for instance, principles that are intended to deal with the structures of injustice (i.e. capitalism, patriarchy, institutional racism).

c) thirdly, the various candidates are advanced on the basis of public reason. Arguments for the difference principle would have to work on different fronts. For instance, citizens, following public reason, would need to establish first whether justice as fairness would pass the test of reciprocal justification. In the (likely) event of many candidates satisfying the reciprocal justification test, other considerations would

³⁴ Van Parijs, P., *Real Freedom for All* (Oxford: Oxford University Press, 1995).

need to be invoked. For instance, whether and how the difference principle does in practice reflect a truly maximin principle, and whether its implicit assumption of continual economic growth, at least in the medium term, might match the economic prospects of the society in question.

d) finally, a selection might be made on the basis of how far each candidate can potentially fulfil the ideal of democratic citizenship. As an example of this, consider the following. Imagine a society with fairly optimistic economic prospects. In b the difference principle is confronted with another serious candidate, for instance, the principle of an unconditional basic income, a principle establishing a tax-free disposable income to be paid to every adult. Entitlement to this income does not depend on the ability/inability to work, therefore it does not distinguish between able/disabled, employed/unemployed, and it does not focus on the family unit. Behind this principle there is an implicit acceptance that certain activities that are indispensable to a society like the rearing of children, or looking after the ill or the elderly, must be economically compensated. Now we need to assess this principle in connection with the difference principle, trying to establish which one is more suited to meet the requirements of citizenship. Of course, it is possible that, given that society's favourable economic circumstances, the difference principle may issue in a substantially higher distribution, and therefore, it might well be preferred. The key point, however, is that when this assessment is carried out, we look at the features of the two (or other) principles, trying to bring out the aspects that connects them more closely to the idea of democratic citizenship. Hence, arguments for the justificatory rationale for the difference principle, or for any other principle of economic justice, will depend on the particular historical circumstances of a society, and must be able to stand the democratic test of justification along the suggested lines.

This account is rather abstract and, to a certain extent, indeterminate. However, a political theory can only hope to uncover the different criteria required in a political justification, through public reason, of principles of justice. In setting out these programmatic criteria, we are implicitly acknowledging that justice as fairness may be a realistic utopia. Of course, whether or not justice as fairness is actually politically justified and adopted as the preferred conception of justice for broadening the constitution is not something that a political theory can establish.

It may well be the case that, in the end, no single conception of justice is adopted, but ‘an orderly contest between them over time is a reliable way to find which (...) is the most reasonable.’³⁵ In this case, political liberalism would have still shown how, through public reason, the difference principle may be adopted.

5.Stability and Legitimacy.

To conclude this chapter I wish to pay a final visit to the idea of legitimacy. As we have seen, the idea of legitimacy, from which the requirement of reciprocal justification derives, is highly demanding. Reciprocal justification applies to both, constitutional essentials and matters of basic justice. However, we have seen how, given pluralism, essentials may be contested, if not in their statement, at least in their further specification. To be sure, the main problem of political liberalism resides in its conflation of three different issues: legitimacy, stability and justification.

Political liberalism is concerned with four types of questions:

1. ‘How is it possible for there to exist over time a just and stable society of free and equal citizens, who remain profoundly divided by reasonable religious, philosophical and moral doctrines?’³⁶

³⁵ PL, p.227.

³⁶ Ibid., p.4.

2. 'How is it possible for citizens of faith to be wholehearted members of a democratic society when they endorse an institutional structure satisfying a liberal political conception of justice with its own intrinsic political ideas and values (...)?'³⁷
3. 'What is the most appropriate conception of justice for specifying the fair terms of social cooperation between citizens regarded as free and equal, and as fully cooperating members of society...?'³⁸
4. How can political liberalism 'uncover the conditions of the possibility of a reasonable public basis of justification on fundamental political questions?'³⁹

The first question suggests a concern with stability; the second, a concern with legitimacy; the third, a concern with justice, and, finally, the fourth, a concern with justification.

In the same way that the questions of political liberalism are not presented with a clear demarcation, each of the answers has not been properly detached from the overall aim of the project. It appears that it is the problem of stability (a problem inherited from TJ) that leads Rawls to the principle of legitimacy. Stability, in his view, requires legitimacy since 'the kind of stability required of justice as fairness is based (...) on its being a liberal political view, one that aims at being acceptable to citizens as reasonable and rational.'⁴⁰ The idea of legitimacy also connects with Rawls's particular conception of democratic political power: that power belongs to the collective body of citizens. For political authority to be legitimate, justice as fairness must be addressed to each citizen's reason. We may say that, through the liberal

³⁷ Ibid., p.xl.

³⁸ Ibid., p.3.

³⁹ Ibid., p.xxi.

⁴⁰ Ibid., p.143.

principle of legitimacy, Rawls presents a highly idealized, and rather demanding, ideal of citizenship where each citizen is considered a constituent part of a collective body.

To be sure, Rawls's theorization of legitimacy differs in an important way from its traditional theorization. Legitimacy usually refers to the question of how, and on what basis, citizens can accept the power of the state. It involves examining the sources of citizens' allegiance to political authority. In relating legitimacy to the criterion of reciprocal justification, Rawls has also made legitimacy unattainable. For, given the fact of pluralism, it is highly unlikely that citizens can achieve that level of agreement, based on unanimity, or quasi-unanimity.

Nothing that I have said here affects the general argument of my thesis, namely, that for justice as fairness to be a realistic utopia, some alterations need to be made with regards to the arguments for its stability and its justification. Issues surrounding legitimacy should be taken up separately, for they are different from the questions of 1) how justice may be justified, given pluralism, and 2) how a constitutional democracy may attain institutional and political stability.

With regards to the question of legitimacy proper, i.e. on what basis may citizens accept the power of a political authority, political liberalism should be concerned with democratic, not liberal, legitimacy. The crucial feature of democratic legitimacy is its sensitivity to diversity. Hence, it acknowledges that there may be different routes to citizens' acceptance of a liberal-democratic order. Some citizens may accept this authority on a liberal basis themselves, on the basis that the liberal state provides the most extensive scheme available of civil rights. A liberal democratic state may also claim citizens' compliance on a democratic basis, on the extent of the political rights that it sanctions. Or, finally, citizens' allegiance to a constitutional

democracy could be founded on prudential grounds. The picture of democratic legitimacy that emerges is, therefore, one of a 'patchwork' of reasons.⁴¹

⁴¹ Evans, M., 'Pluralising Liberalism, Liberalising Pluralism', *Res Publica*, 10 (2004), pp.449-60, p.460.

CHAPTER 7

OVERLAPPING CONSENSUS REVISITED

This thesis has argued that for justice as fairness to be possible, the project of political liberalism needs to dispense with the idea of overlapping consensus. This conclusion has been reached after examining a crucial feature of constitutional democracies, namely, the fact of reasonable pluralism. However, I am prepared to give overlapping consensus, as it were, a one last chance. Hence, in this concluding chapter, I survey four ways in which this idea may be preserved. In particular, I discuss the possibilities of overlapping consensus in terms of it being a realistic utopia, an unrealistic utopia, a dystopia, or an account of critical moral justification. My general aim is to assess whether any of these interpretations is somehow able to rehabilitate overlapping consensus and to place it back at the centre of the project of political liberalism. Such an assessment must be carried out against the objective of moral stability. We have also established that moral stability is attained when the majority of the active citizens of a particular society have internalised the political values that are embedded in their society's basic institutions and have incorporated them into their respective domains of value. That moral stability should be our concern here must be clear from the argument sustained throughout this thesis. If overlapping consensus is required neither for institutional nor for political stability, and if the fact of pluralism demands that justice be justified on a political basis alone, all we can do is to look at whether overlapping consensus can, and should, be saved as a requisite of moral stability.

1.Overlapping Consensus: realistic or unrealistic utopia?

As I mentioned in my Introduction to this thesis, a project of political philosophy can be regarded as a realistic utopia if, by stretching the possibilities of actual feasibility, it fulfils the key role of reconciling ourselves to the political and social conditions of our times¹. In the so-called ‘domestic’ context, which generally refers to more or less well-established constitutional democracies, the limits of practical possibilities are defined by the fact of reasonable pluralism. Limiting what may be practically possible must not imply that we are subject to a kind of ‘dictatorship of the actual’. Political and social change may be advanced and the role of political philosophy is to hypothesize about the possibilities of ideal theory. Before assessing the extent to which overlapping consensus can be considered a realistic utopia, let us look at the features that, in Rawls’s own terms, a realistic utopia ought to exhibit².

According to Rawls, there are seven conditions that a political conception of justice must fulfil for these purposes. Two of these conditions, if met, will make such a conception ‘realistic’. Firstly, a conception of justice can be said to be realistic if it achieves stability ‘for the right reasons’, where such stability does not rest on a conception of human nature and of ideal political society that is too far removed from what we know is, or may be, possible. In other words, the condition of realism requires that our account of human nature and of ideal political society correspond to the ‘laws of nature’³, for there is no point in imagining a world that cannot be. The conjecture here is that citizens may act in accordance with the principles of justice that they have acquired from their experiential knowledge of living in a just society, and

¹ See LoP, p.11 and IPPR, pp.3-4.

² Here I follow Rawls’s definition in LoP, p.11.

³ Ibid., pp.12-13.

this affords stability for the right reasons. However, in accounting for this possibility, we have not transcended the limits of what is, or may be, feasible.⁴

The second condition for 'realism' is that a conception's principles are workable when applied to a specific political and social order. Rawls exemplifies this by comparing the simplicity in the application of his primary goods with the difficulties of applying a utility principle, or Sen's basic capabilities principle. Being workable seems to require that our principles must require the minimum amount of information so that applying them is a more or less straightforward matter.⁵

There are five conditions that a conception of justice must fulfil for it to be (realistically) 'utopian'. Firstly, a political conception is realistically utopian when it proposes political ideas, principles and concepts and organises them into an account of a reasonable just society. There may be many reasonable conceptions of the just society, all of which must present the following features: 1)it must specify basic constitutional rights and principles, 2)it must prioritise those rights, liberties and opportunities over other ideals and principles, and 3)it must provide adequate means for the purposeful use of basic rights and liberties⁶. These conceptions should also offer a requirement of reciprocity so that reasonable political proposals may be judged against this standard. Political proposals based on reasonable conceptions must be such that those proposing them can, in principle, expect others to endorse them as free and equal citizens. But, principles will require some interpretation and this is where the variety of reasonable liberal conceptions may come into play. Each of these

⁴ Rawls explicitly quotes an opening section of Rousseau's *Social Contract*: 'My purpose is to consider if, in political society, there can be any legitimate and sure principle of government, taking men as they are and laws as they might be.' *Ibid.*, p.13.

⁵ That principles need to exhibit a certain simplicity in their application seems a constant concern of Rawls's. The reader might recall from chapter 3 that he makes this point in chapter 8 of TJ when discussing the content of the sense of justice associated with justice as fairness; he suggests that this simplicity of application contributes towards a greater relative stability than the one afforded by the principle of utility. The requirement of 'simplicity' is also a feature of the constitutional essentials.

⁶ *Ibid.*, p.14.

conceptions offers a particular ordering and balancing of principles and values and, hence, each is guided by different public reasons. Therefore, to be utopian, a conception of justice must belong to the general class of liberal conceptions.

Secondly, a realistically utopian conception must be political in character, that is, it must exclusively incorporate political ideas. Such ideas are certainly moral ideas, but they must be taken from the fund of political ideas present in the public culture of democracy and, more importantly, they must be taken as such without further reference to comprehensive ideas⁷. A utopian conception stays within the political in both its theorization and application. For instance, the idea of the person as free and equal defines a political conception of citizenship, and not a metaphysical view of what persons are.

Thirdly, a realistically utopian conception must offer some account of necessary political virtues and of how they can be realised. The political virtues ('a sense of fairness and tolerance and a willingness to meet others halfway'⁸) are nurtured by the political and social institutions of a constitutional democracy. Citizens brought up within such institutions have acquired a kind of 'internal' knowledge of the principles and ideas of the political conception so that it is 'natural' for them to exhibit compliant behaviour, and therefore to contribute to stability for the right reasons. General (moral) compliance contributes to a just and stable political order for 'the idea of realistic utopia is importantly institutional'⁹.

Fourthly, since social unity, given pluralism, cannot rest on religious, philosophical or moral unity, and if we are going to go beyond a *modus vivendi*, a realistically utopian conception must be accepted on the basis of an overlapping consensus of comprehensive doctrines.

⁷ Ibid., p.15.

⁸ Ibid.

⁹ Ibid., p.16.

And fifthly, a conception must offer a political view of toleration in order to act as a guarantee in the cases where comprehensive doctrines lack one. A political conception is strengthened if it is able to provide a view of toleration based on public reason.

Having outlined Rawls's idea of a realistic utopia, we must now assess whether, in the light of these characteristics, the idea of overlapping consensus belongs to such an account. To be sure, the term 'utopia' refers to a conjectural view of the 'best possible', but 'not yet attainable'. Our question is, therefore, to elucidate whether overlapping consensus is the 'best possible', but 'not yet attainable', account of moral stability. For the sake of clarity, let us recall here the two requirements of moral stability: 1) citizens' internalisation of the principles and values that ground their institutions and 2) the incorporation of such principles and values into citizens' overall moral perspectives. It might seem relatively straightforward to see how overlapping consensus does not fulfil the 'not yet attainable' requirement. As we have argued in this thesis, the fact of pluralism is a very obstinate fact and one that does not sit comfortably with the idea of a homogeneity of perspectives. However, Rawls asks us to stretch the possibilities of actual feasibility in order to foresee how we may be reconciled with the conditions of our world. But to this we might ask: can we foresee the circumstances under which our plural societies might obtain the similarity of views that is required for overlapping consensus? Three reasons can be shown that dictate against this possibility:

i) the fact of pluralism understood primarily as a democratic fact. Here we assume, from a realistic utopian position, that our political future will continue to be based around the idea of democracy, however it might be politically realised. The realistic hope is that future citizens will continue to support institutions that further liberal

principles and democratic institutions, and to advance those wherever they are not available. Whatever new political institutions are required, for instance to deal with the forces of globalisation, they must conform to, at least, a minimally democratic benchmark. In such realistically utopian circumstances, the scope of pluralism is bound to increase, not decrease. As different cultures, religions, philosophical and political traditions come into increasing contact with one another, the extent of diversity will become more evident, particularly at local level. It would be futile to ask whether this would be a good or a bad thing. The important aspect here is that pluralism is both a democratic grounding and a democratic result, for it seems that the stronger the desire of being politically heard, the stronger the motivation for bringing about, or strengthening democracy. Hence, even though some people may still relish the thought of living in a relatively homogeneous and self-contained society, our political future seems to be pointing in a different direction. Those societies that stubbornly remain homogeneous –like North-Korea or Cuba, for instance- are hardly examples of the good society. Thus, the social and political conditions of our world make radical pluralism, to a large extent, consubstantial with democracy.

It might be argued that there is nothing consubstantial between pluralism and democracy, and that the link is, in fact, contingent and restricted to the circumscribed historical period of Modernity (and Post-Modernity). After all, Greek society, for instance, can hardly be named as an example of a diverse and heterogeneous society. But this objection misses the point. I am not saying that a society that is homogeneous in its social, cultural, religious and political manifestations (for instance, Cuba) necessarily falls short of a democratic standard, only that democratic societies are

unlikely to remain homogenous. And this is because, given our current predicament¹⁰, human activity under free conditions leads to pluralism, and pluralism both grounds and requires democracy. There may well be cultural reasons behind the social and cultural homogeneity of a society like Cuba, and political philosophy would have little to say about this. However, there are certainly political reasons for its political homogeneity, and they are the result of the fact of oppression. Of course, Rawls does acknowledge this, but he treats pluralism as simply something that has to be recognized for political practical purposes, rather than as a force for change and democratisation.

ii) the burdens of judgement. We can clearly see here the extent to which Rawls has underestimated the importance of this notion. As I have argued in this thesis, it creates a fatal difficulty for his idea of overlapping consensus. From the preceding discussion, we can conclude that the burdens of judgement will continue to hold against the possibility of a highly moralized agreement. This is not just because of the extent of pluralism itself, however important this may be. As societies become more interconnected and the domestic context of states is more comprehensively incorporated into a variety of international and global networks, the range of political issues and their complexity are bound to expand, and hence, the evidence, theories, principles and values from which to make binding political decisions. In the case of already democratised societies, issues of risk under uncertainty¹¹, which do not belong to the set of 'traditional' political matters, may add to the centrality of disagreement. Under these conditions, the burdens of judgement appear to cut rather strongly against the feasibility of overlapping consensus. Furthermore, issues of institutional design

¹⁰ By this I mean the extent of technological development and, perhaps even more crucially, the opportunities to access information.

¹¹ See, for instance, Beck, U., *The Risk Society* (London: Sage, 1992), and *Ecological Politics in an Age of Risk* (Cambridge: Polity, 1995).

may need to be addressed at some point. For instance, in the last few years, we have seen the development within political philosophy of theories of deliberative democracy. When thinking about the extent and limits of deliberative democracy, one of my concerns has always been the issue of the practical institutionalisation of this brand of democracy. It may well be the case that the traditional channels of policies and political opinions such as parties or unions may be surpassed, and other channels may be established. This might entail further changes in the political process of democracy as we currently know it. It may well be said against this that Rawls's idea of public reason still grants us the possibility of participating in the possible processes of discussion and decision-making on these issues. However, as I argued in chapter 4, the extent of political pluralism -that would seem central in the discussion of this type of issues- is rather tamed within an overlapping consensus.

iii)the fact of oppression. The final reason why an overlapping consensus may not be realistic is a crucial one for, if pluralism is consubstantial with democracy, it seems that the only way in which moral stability can be achieved is by using state power or the power of any other relevant political authority. Rawls himself would not sanction the kind of stability that results 'for the wrong reasons'. There are two problems here. One is that it is very difficult to 'measure' the degree of citizens' internalisation of political and social values, for internalisation may not result in any particular type of external conduct on citizens' part. Citizens may go about their everyday businesses having, more or less willingly, accepted to being governed by a particular political authority (legitimacy), but not having internalised the appropriate set of political values. 'Internalisation' is a rather intimate matter, and therefore not a political issue. This leads us to the second problem. Since whether or not citizens ultimately endorse political values 'all the way down' is not a suitable matter for politics, it is covered by

the basic personal freedoms. Justice as fairness itself would cover citizens from possible interference by the state or by a relevant political authority in this respect. It is interesting to note that this aspect that makes overlapping consensus unrealistic, also makes it non-utopian.

So far we have shown that overlapping consensus is not realistic in Rawls's own terms. It is, therefore, not a realistic utopia. But, is it an unrealistic utopia¹²? In other words, wouldn't it be a good thing for societies to exhibit the degree of similarity between people's moral views that is required for overlapping consensus?

The answer to this question from a utopian perspective must be affirmative. For the utopian mind, the vision of a society whose basis of social unity rests on the coherentism of moral perspectives may appear rather attractive. This vision presents an alternative to our current predicament, an alternative to the subjectivism, scepticism and relativism that are rife in modern societies, and one that is morally superior to our present condition.

Is there anything to be gained by living in an arcadia of highly 'moralized' selves? The answer to this question will depend on how well one's views sit within the conception of utopia itself. But if utopia aims to transcend a given social and political order, rather than to change it, and if it aspires to 'perfection' (however defined), we may be led to reject such a utopian vision. Certainly, a society whose basis of social unity is supplied by an overlapping consensus seems to exhibit these two traits, for it supposes a vision of the perfect society as one in which citizens' political and moral outlooks thoroughly cohere, and, as I have tried to show in this thesis, it appears to bypass pluralism, rather than to engage with it.

¹² To clarify: by utopian here I mean 'unrealistically utopian'. Rawls would regard this as utopian in the pejorative sense.

Perhaps the most important question is not whether we gain anything by living in utopia, but whether we lose anything. Here we might say that we clearly stand to lose if political philosophy continues to concern itself with questions that may be outside of our present predicament, questions that are too far removed from the lives of citizens. Hence, whereas it may be a good thing to live in a society that has attained the goal of overlapping consensus if our aspiration is to become saints or heroes, the citizens of that society would not have much to say about politics. And, where there is a society, a whole host of political matters (scarcity, distribution, entitlements, to name but a few 'classical' matters) necessarily arise. Hence, where there is a society, there is politics and overlapping consensus appears to obstruct our ability to fully engage in its exercise. To put it in the words of Thomas Nagel: 'A theory is utopian in the pejorative sense if it describes a form of collective life that humans, or most humans, could not lead and could not come to be able to lead through any feasible process of social and mental development. It may have value as a possibility for a few people, or as an admirable but unattainable ideal for others. *But it cannot be offered as a general solution to the main question of political theory: How should we live together in society?*'¹³

2.Overlapping Consensus: A Dystopia?

A theory or view about the present can project itself into the future by exhibiting its best or its worst features¹⁴. Behind this is the idea that failing to bring about a utopia may mean that we might, in the end, come to live in a dystopia. For our purposes here we need to discuss whether the alternative to a (utopian) society that has attained an overlapping consensus is such a bad thing after all.

¹³ Nagel, T., *Equality and Partiality* (New York: Oxford University Press, 1991), emphasis added.

¹⁴ Crook, S., 'Utopia and Dystopia' in Browning, G., Halcli, A., and Webster, F. (eds.), *Understanding Contemporary Society* (London: Sage, 2000), pp.205-218, p.205.

The alternative to the overlapping consensus society would be a society marked by the fact of radical pluralism and the dystopian claim will thus say that living in that kind of society would be a good thing because we will gain some valuable diversity among human beings. Pluralism, in the dystopian view, is not a regrettable fact and needs to be celebrated in its true extent.

First of all, I would need to explain what I mean by radical pluralism. The term 'radical' might invite misunderstanding, but my intention here is to relate it to its etymological meaning. Hence we may think of the 'roots' of pluralism in the sense of the 'sources' of pluralism. The important point is that liberal democracies usually contain multiple sources of pluralism, although as it may be expected, some will be more relevant than others. In my view, one of the main problems of the Rawlsian project is the narrowness with which it, not just defines pluralism, but also conceives of its extent. A pluralism of religious, philosophical and moral doctrines reflects only a small portion of a society's views and positions.

It is important that we do not confuse the 'sources' of pluralism with something like the burdens of judgement. Whilst the burdens of judgement mainly focus on the process of reasoning as such (they are theorized, after all, as the sources of reasonable disagreement), an account of the sources of pluralism needs to focus on any aspect of human activity that in itself might produce a variety of options and viewpoints. This certainly covers religious, philosophical and moral doctrines, but also economic theories, aesthetic judgements and ethical convictions. Radical pluralism assumes that human activity (and not just human reason) under free conditions normally results in a variety of positions (and not just doctrines). Hence, most spheres of human activity, political, social, religious, economic, scientific, familial, etc. will issue in a plurality of positions.

This version of pluralism is, in my view, more closely connected to the ideal of democracy than Rawls's version. I said above that pluralism both grounds and requires democracy. It requires democracy in the simple sense that, whenever there is a variety of points of view at stake, democracy steps in to adjudicate. However, pluralism also grounds democracy through an ideal of (radical) democratic citizenship. This idea, like Rawls's idea of citizenship, is also underpinned by a general understanding of freedom and equality. Therefore, citizens may make claims on the state, and on one another, on this basis. However, the values of freedom and equality, although free-standing and political, are also general. Citizens may hold a variety of interpretations of these values and their requirements, with the important criterion here being one of relevance. Hence, instead of asking citizens to embed their political values into their own general doctrines, the view from radical pluralism asks citizens to exercise their individual judgement with regards to how relevant the reasons and values that they wish to advance in politics are. Take the example of a Marxist, for instance. In Rawls's view, a follower of Marx's economic and political theories would either be unreasonable (if, for example, politically advocating, from his own comprehensive doctrine, policies that may advance the common ownership of the means of production), or would have to convince herself that, all things considered from within her own Marxist doctrine, the difference principle is the best way of realizing economic justice. This clearly mitigates the extent of pluralism. By contrast, radical pluralism would allow Marxists to politically advocate policies that may advance a communist society with the only condition that those policies be based on relevant reasons, in this case on relevant political, social and/or economic reasons. Hence, arguments from, for instance, the benefits of common ownership on, say, the environment, would be allowed. However, arguments based on the necessity to

overthrow capitalism in order to bring about the dictatorship of the proletariat would not have a place within relevant political argument as they belong to the internal theoretical construction of Marxism.

Would this society be preferable to an overlapping consensus society? The utopia claim would prefer the orderly life of a society where citizens' views are highly harmonized, whereas the dystopia claim would prefer the 'messy' life of radical pluralism.

One final point on Rawls's narrow conception of pluralism. Political liberalism reflects a view of citizens as 'neat' selves that possess two kinds of views (a political and a general moral view) that can be potentially harmonized without major difficulties. This view does not sit easily with the rather fluid and kaleidoscopic identities that citizens exhibit in contemporary societies. Political liberalism endorses a highly compartmental view of pluralism, where the salient features of citizens' personal identities are the religious, philosophical and moral perspectives that they affirm. By making comprehensive doctrines the only relevant type of pluralism to be considered for the purposes of political philosophy, political liberalism has contained pluralism and situated it within the realm of highly speculative doctrinal conflict. This conception of pluralism which, to a certain extent albeit in a different way, is shared generally by multiculturalists, is not only narrow, but also rather 'essentialist', as it places people into pre-defined categories that have already been conferred a politically relevant status. By contrast, radical pluralism acknowledges a variety of citizenship positions whereby people prioritise the parts of their political identity that may be relevant to the different political contexts in which they act politically. It is then up to citizens, both individually and collectively, to decide where a particular feature of their identity may be politically relevant, and when it may be so. From the dystopian

perspective, therefore, citizens may not be ‘orderly’ selves politically and morally, but we need not lament this because there is too much to gain by embracing radical pluralism.

3.Overlapping Consensus and Critical Justification.

This final section of my thesis is concerned with an examination of Rawls’s idea of overlapping consensus as an account of the critical justification of justice. I will firstly examine Rawls’s own approach to critical justification. Here my main aim is to highlight the problems of an approach that takes comprehensive doctrines as the basis for the critical justification of justice. I then conclude that we should follow an alternative path. This alternative path directs the question of the critical justification of justice not to doctrines, but to citizens. However, even in this alternative version, overlapping consensus supersedes the boundaries of political theory, and it must, therefore, be regarded as an idea that belongs to the province of moral theory.

But, why should we concern ourselves with critical justification? My answer to this is that we are led directly to critical justification once we realize that moral stability is something that cannot be addressed by the normative theory itself. As I have argued, the normative theory of political liberalism is only able to provide relatively solid arguments for institutional and political stability. Hence, critical justification is not part of the normative theory itself. It is, rather, external to it. It might be asked then that, if critical justification is external to the theory, there is not much to gain from it. To this I would claim that, in Rawls’s own terms, there could be a welcomed increase in moral stability. But looking into critical justification does not constitute an argument for stability. It is rather like the argument for congruence in TJ, which could only show how a society well-ordered by justice as fairness *could* look

like. I emphasize 'could' because this thesis has shown that the well-ordered society of justice as fairness does not require moral stability.

Let us first consider critical justification from a standard Rawlsian point of view. A particular conception of justice may become critically justified if citizens generally accept that conception in the light of principles, beliefs, values and ideas that are readily available to them. The way Rawls approaches this question in PL is by looking at how the different comprehensive doctrines relate to his preferred political conception.¹⁵ The purpose of the last section of Lecture IV is to show that the particular realm of values of the various reasonable comprehensive doctrines either support, are congruent with, or at least do not conflict with, the political values that underlie the political conception of justice. If Rawls can show that this is not just a mere possibility, but a very likely one, the political conception will, as a result, have been critically justified. To accomplish this task, he uses a multipolar strategy that takes Kantianism, Utilitarianism, the religious view with an account of free faith and the pluralist view of values as model cases of the set of possible philosophical grounds for the political conception.¹⁶

Consider the following statements about the kind of relationships Rawls has in mind:

- 1.The political conception of justice and its principles of justice can be derived via deduction from the Kantian moral idea of autonomy.
- 2.Utilitarians who follow Bentham and Sidwick (but possibly utilitarians of other persuasions as well) will eventually come to accept the political conception as the one

¹⁵ In looking into this matter, I follow Rex Martin's suggestion that Rawls characteristically describes critical moral justification as being one of overlapping consensus; see *System of Rights* (Oxford: Clarendon, 1993), ch.13, p.334.

¹⁶ PL, pp.77-78, 168-72, 144-50.

that best approximates to the principle of utility given the conditions of knowledge present in modern societies marked by the fact of uncertainty.

3. The comprehensive pluralist view already contains the political values within its wider realm. The political conception can be endorsed by a balancing of judgements that prioritises the political over the non-political values.

4. A religious doctrine with an account of free faith will accept the political conception on the basis of the principle of toleration since it underpins the fundamental basis (the basic liberties) of a constitutional regime.

In each of the four cases, the political conception is said to be somehow related to the particular comprehensive doctrine. However, it is not clear what exactly needs to be embedded into a comprehensive doctrine. We have three possibilities: justice as fairness as a whole, the two principles of justice, or any political conception of a broadly liberal character. We have therefore a minimum (general liberal conceptions) and a maximum (justice as fairness as a whole) programme for the enterprise of critical justification. For the maximum programme, the set of comprehensive doctrines would all need to embed: i) the centrality of the political values and of public reason, ii) equal basic rights and liberties specified by the first principle, iii) fair equality of opportunity; and iv) the difference principle. For the critical justification of the two principles ii-iv will be required, and only ii for the minimum programme.

Let us concentrate on the least demanding of all cases. This is the case where the pluralist view of values embeds the values that support a general understanding of liberal justice. First of all, we need to see what Rawls says about the pluralist view. The main feature of this view is that it is not completely articulated; it contains political values that belong to a free-standing conception of justice and some non-political values where 'each subpart of this family has its own account based on ideas

drawn within it, leaving all values to be balanced against one another, either in groups or singly, in particular kinds of cases.’¹⁷ In the pluralist view, each domain of value is free-standingly endorsed, and the political conception is affirmed on the strength of the pluralist view’s political realm of values. Judgements made from within the pluralist view normally support political values over other values that may oppose them.¹⁸

It is not difficult to see how the pluralist view could support general liberal conceptions of justice. However, it is far more difficult to ascertain whether the balancing of political values, even granting their overridingness over non-political ones, would always issue in particular types of judgements in support of a liberal conception of justice. Being pluralist and political does not guarantee us the required result, i.e. a liberal political conception. Rawls seems to have assumed that political views that are non-comprehensive are necessarily liberal. However, politically plural judgements may, for instance, support a libertarian balancing of values.

Part of the problem with the pluralist view is that its theoretical articulation is rather loose. This seems to have been deliberate on Rawls’s part as he uses this view to explain how a *modus vivendi* might turn into a constitutional consensus.¹⁹ However, we are unclear about the specific theoretical resources at our disposal with which to effect a thorough concordance of fit.

Let us now consider the example of a doctrine with a conception of rationality thicker than justice as fairness, like Catholicism, and let us see how it can embed the two principles so that they may become critically justified by Catholics. Let us focus, in particular, on how Catholicism may effect a concordance of fit between the difference principle and the rest of their morality. The first point that we would need

¹⁷ Ibid., p.145.

¹⁸ Ibid., p.170.

¹⁹ Ibid., pp.159-60.

to ascertain here is whether there is a particular dimension of Catholicism that would be relevant for our purposes. There are various possibilities here. We could start from the Gospel, i.e., the parable of the good Samaritan, from classical doctrinal corpuses, i.e. Aquinas, or from more or less recent Papal Encyclicals, i.e. *Populorum Progressio* or *Dignitatis Humanae*. It may be said that, given the philosophical nature of the exercise, we should start with the latter. On precisely this basis, Leif Wenar has shown, rather persuasively, how Catholicism could support the two principles, although it could not equally support the burdens of judgement and the idea of reasonableness.²⁰ On this account the two principles could be said to be critically justified to Catholics. However, the problem is that, even though Catholicism as a body of thought could embed the difference principle, this does not entail that Catholics themselves are able to do the same, for the simple reason that the exercise has taken place at too abstract a level and on the basis of reasons that are not usually available to the average Catholic believer.

Another option would be to take the Gospel parable as our starting point, for most Catholics are generally familiar with Gospel stories.²¹ The problem here is that general Gospel stories cannot embed fairly specific political values one way or the other. Concern for strangers does not necessarily issue on anything as particular as the difference principle, or even distributive justice.

Rawls argues that from a general aspect of Catholic social doctrine, such as the special concern for the poor, Catholicism could embed the difference principle, or some similar idea.²² But this would be the case if the difference principle, or an analogous one, were the only principle of justice available in society.

²⁰ 'An Internal Critique'.

²¹ Rawls himself uses this example to illustrate his point about the wide view of public culture, IPRR, p.594.

²² *Ibid.*, n55.

From this we may conclude that the problem really is whether reasonable comprehensive doctrines have the resources to deal with the process of critical justification. As we have seen, there are some difficulties here, and they arise from a kind of philosophical and moral indeterminacy of overlapping consensus.

Hence, critical justification might proceed in accordance with an alternative approach. This approach directs the process of critical justification to citizens themselves. Here the process itself is not mediated by the comprehensive doctrines that citizens affirm. It is up to citizens themselves to make their own connections between their political and their non-political values. Hence, this would be an open-ended, fluid and continuous process whose flows would not only go from the moral to the political perspective, but vice versa.

Since this is a process for citizens to effect, its results are not pre-defined, nor its possibilities advanced. This view reflects a much thinner version of the connection between citizens' different domains of value, which differs in important ways from the aim of overlapping consensus. But even if we direct the question of coherence to citizens, rather than doctrines, the exercise turns out to be a moral, and not a political, enterprise. All that a political theory can do is to specify the conditions for an ethos of citizenship, for it is neither unreasonable nor illiberal for citizens to develop a principled bifurcation between their political moralities and the rest of their values. And, as I have argued throughout this thesis, there are reasons, internal to the project of political liberalism itself, that recommend that we stay within the political. Only from a political perspective, and not from the realm of moral theory, are we able to claim that justice as fairness may be a realistic utopia after all.

CONCLUSION

I started my thesis by suggesting that if justice as fairness is to be regarded as a realistic utopia, some alterations need to be made to the argument about the possibilities of its realization offered by Rawls. I also suggested a working conjecture, namely, that we need overlapping consensus neither to stabilize justice as fairness, nor to justify it. And, I also proposed to examine stability from three points of view: institutional, political and moral.

I have shown how a project of political theory might be concerned with the first two kinds of stability –institutional and political-, but also, how it might become theoretically burdened by introducing arguments from moral stability. The main reason for this is that the fact of pluralism makes it, not just implausible, but also undesirable. Rawls's account of reasonable pluralism also required some revision in order to reflect more accurately the true extent of pluralism in contemporary societies. The important consequence that I extracted from those two conceptual revisions (stability and reasonable pluralism) was that, in order to preserve institutional stability under the conditions of pluralism, we need to render the stability of institutions separately from the coherence of citizens' overall moralities.

Our concern with institutional stability directed us towards constitutional consensus, an idea introduced by Rawls precisely to show the prospects for justice as fairness. I argued that, although constitutional consensus is necessary for institutional stability, it is not yet a sufficient condition. My argument for the necessity of constitutional consensus was also an argument against the sufficiency of *modus vivendi*.

I then tried to address the question of whether institutional stability was put under strain by adding to the constitution. This seems to have been one of Rawls's main motivations behind overlapping consensus. Whereas Rawls's answer to this stability problem is to establish a certain kind of moral justification (overlapping consensus), I argued that solving stability, beyond constitutional consensus, only requires political justification.

My final step was to examine overlapping consensus in the full knowledge that it was unable to address the relevant problems of stability. I tried to see whether anything could be preserved for the purposes of a theory of politics. I concluded that overlapping consensus is part of moral, not political, justification and, as such, not required for a political theory.

I have ventured into this research project with the main purpose of improving on the Rawlsian framework. Since, according to Rawls, one of the tasks of political philosophy is to stretch the limits of the possible, I have tested the idea of overlapping consensus against his very own standard. My concern, therefore, lays with the overall soundness of Rawls's latter work. The general conclusion of my investigation is that there are important reasons, and reasons that are internal to political liberalism itself, that make overlapping consensus inadequate for a project of political justification.

BIBLIOGRAPHY

Relevant Works by John Rawls

A Theory of Justice (Oxford: Oxford University Press, 1972)

'Justice as Fairness: Political not Metaphysical', *Philosophy and Public Affairs*, 14 (1985), pp.223-51, reprinted in *Collected Papers*, edited by Samuel Freeman (Cambridge, MA: Harvard University Press, 1999), pp.388-414.

'The Idea of an Overlapping Consensus', *Oxford Journal for Legal Studies*, 7 (1987), pp.1-25, reprinted in the *Collected Papers*, pp.421-48.

'The Domain of the Political and Overlapping Consensus', *New York University Law Review*, 64 (1989), pp.233-55, reprinted in the *Collected Papers*, pp.473-496.

Political Liberalism (New York: Columbia University Press, 1993); the revised paperback edition of 1996 includes a new Introduction and the 1995 paper 'Reply to Habermas' as Lecture IX.

'The Law of Peoples', in *On Human Rights: The Oxford Amnesty Lectures, 1993*, edited by Steven Shute and Susan Hurley (New York: Basic Books, 1993).

'Reply to Habermas', *The Journal of Philosophy*, 93 (1995), pp.132-80.

'The Idea of Public Reason Revisited', *The University of Chicago Law Review*, 64 (1997), pp.765-807, reprinted in the *Collected Papers*, pp.573-615.

Collected Papers, edited by Samuel Freeman (Cambridge, MA: Harvard University Press, 1999).

The Law of Peoples with 'The Idea of Public Reason Revisited' (Cambridge, MA: Harvard University Press, 1999).

Justice as Fairness: A Restatement, edited by Erin Kelly (Cambridge, MA: Harvard University Press, 2001)

Anthologies and Collections on Rawls

Arneson, R., 'Symposium on Rawlsian Theory of Justice: Recent Developments', *Ethics*, 99 (1989), pp.695-944.

Corlett, J.A. (ed.), *Equality and Liberty: Analyzing Rawls and Nozick* (London: McMillan, 1991).

Daniels, N. (ed.), *Reading Rawls* (New York: Basic Books, 1975).

Griffin, S. and Solum, L., 'Symposium on John Rawls's *Political Liberalism*', *Chicago-Kent Law Review*, 69 (1994), pp.549-842.

Richardson, H., and Weithman, P. (eds.), *The Philosophy of Rawls: A Collection of Essays*, 5 volumes (New York: Garland, 1999)

Volume I: *Development and Main Outlines of Rawls's Theory of Justice.*

Volume II: *The Two Principles and Their Justification.*

Volume III: *Opponents and Implications of A Theory of Justice.*

Volume IV: *Moral Psychology and Community.*

Volume V: *Reasonable Pluralism.*

Stayn, S. (ed.), 'Symposium on *Political Liberalism*', *Columbia Law Review*, 94 (1994), pp.1813-1949.

Freeman, S. (ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003).

Other Relevant books and articles

Ackerman, B., 'Political Liberalisms', *The Journal of Philosophy*, 91 (1994), pp.364-86.

Audard, C., 'Political Liberalism, Secular Republicanism: two answers to the challenge of pluralism', in *Philosophy and Pluralism*, edited by D. Archard (Cambridge: Cambridge University Press, 1996)

Baehr, A.R., 'Toward a New Feminist Liberalism: Okin, Rawls and Habermas', *Hypatia: A Journal of Feminist Philosophy*, 11 (1996), pp.49-66.

Baier, K., 'Justice and the Aims of Political Philosophy', *Ethics*, 99 (1989), pp. 771-90.

Barry, B., *The Liberal Theory of Justice. A Critical Examination of the Principal Doctrines in A Theory of Justice by John Rawls* (Oxford: Clarendon Press, 1973)

Barry, B., 'John Rawls and the Search for Stability', *Ethics*, 105 (1995), pp. 874-915.

Bates, S., 'The Motivation To Be Just', *Ethics*, 85 (1974), pp.1-17.

Baumeister, A. T., *Liberalism and the 'Politics of Difference'* (Edinburgh: Edinburgh University Press, 2000)

Baynes, K., *The Normative Grounds of Social Criticism: Kant, Rawls and Habermas* (Albany: State University of New York Press, 1992).

Bellamy, R., *Liberalism and Pluralism* (London: Routledge, 1999)

Bellamy, R., and Hollis, M., 'Liberal Justice: Political not Metaphysical', *The Philosophical Quarterly*, 45 (1995), pp.1-19.

Bohman, J., 'Public Reason and Cultural Pluralism. Political Liberalism and the Problem of Moral Conflict', *Political Theory*, 23 (1995), pp.253-79.

Brower, B., 'The Limits of Public Reason', *The Journal of Philosophy*, 91 (1994), pp.5-26.

Caney, S., 'Anti-perfectionism and Rawlsian Liberalism', *Political Studies*, 43 (1995), pp.248-64.

Clarke, S., 'Contractarianism, Liberal Neutrality and Epistemology', *Political Studies*, 47 (1999), pp.627-42.

Cohen, J., 'Moral Pluralism and Political Consensus' in *The Idea of Democracy*, edited by Copp, D., Hampton, J. and Roemer, J.E. (New York: Cambridge University Press, 1993), pp. 270-91.

Cohen, J., 'Pluralism and Proceduralism', *Chicago-Kent Law Review*, 69 (1994), pp. 589-618.

Cohen, J., 'Taking People as They Are?', *Philosophy and Public Affairs*, 30 (2002), pp. 363-86.

Copp, D., 'Pluralism and Stability in Liberal Theory', *The Journal of Political Philosophy*, 4 (1996), pp. 191-206.

Daniels, N., *Justice and Justification. Reflective Equilibrium in Theory and Practice* (Cambridge: Cambridge University Press, 1996).

Doppelt, G., 'Is Rawls's Kantian Liberalism Coherent and Defensible?', *Ethics*, 99 (1989), pp.815-51.

Estlund, D., 'The Survival of Egalitarian Justice in John Rawls's *Political Liberalism*', *The Journal of Political Philosophy*, 4 (1996), pp. 68-78.

Estlund, D., 'The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth', *Ethics*, 108 (1998), pp. 252-75.

- Evans, M., 'Is Public Justification Central to Liberalism?', *Journal of Political Ideologies*, 4 (1999), pp.117-36.
- Evans, M., 'Public Reason as Liberal Myth', *Journal of Transatlantic Studies*, 1 (2003), pp.8-25.
- Evans, M., 'Pluralising Liberalism, Liberalising Pluralism', *Res Publica*, 10 (2004), pp. 449-60.
- Exdell, J., 'Feminism, Fundamentalism and Liberal Legitimacy', *Canadian Journal of Philosophy*, 24 (1994), pp. 441-64.
- Frazer, E., and Lacey, N., 'Politics and the Public in Rawls' Political Liberalism', *Political Studies*, 43 (1995), pp. 233-47.
- Freeman, S., 'Reason and Agreement in Social Contract Views', *Philosophy & Public Affairs*, 19 (1990), pp. 122-57.
- Freeman, S., 'Political Liberalism and the Possibility of a Just Democratic Constitution', *Chicago-Kent Law Review*, 69 (1994), pp. 619-68.
- Fulliwinder, R.K., 'Citizenship, individualism and democratic politics', *Ethics*, 105 (1995), pp.497-515.
- Galston, W.A., *Liberal Purposes* (Cambridge: Cambridge University Press, 1991).
- Gaus, G.F., 'Reasonable Pluralism and the Domain of the Political: How the Weaknesses of John Rawls's Political Liberalism Can Be Overcome by a Justificatory Liberalism', *Inquiry*, 42 (1999), pp.259-84.
- Greenawalt, K., 'On Public Reason', *Chicago-Kent Law Review*, 69 (1994), pp. 669-89.
- Griffin, S.M., 'Political Philosophy Versus Political Theory: The Case of Rawls', *Chicago-Kent Law Review*, 69 (1994), pp. 691-707.
- Gutmann, A. and Thompson, D., 'Deliberative Democracy Beyond Process', *The Journal of Political Philosophy*, 10 (2002), pp. 153-174.
- Gutmann, A., 'Rawls on the Relationship between Liberalism and Democracy', in Freeman, S. (ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003).
- Habermas, J., 'Reconciliation Through the Public Use of Reason: Remarks on John Rawls's Political Liberalism', *The Journal of Philosophy*, 92 (1995), pp. 109-31.
- Hampton, J., 'Should Political Philosophy Be Done Without Metaphysics?', *Ethics*, 99 (1989), pp.791-814.

- Harsanyi, J.C., 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory', *American Political Science Review*, 69 (1975), 594-606.
- Hart, H.L.A., 'Rawls on Liberty and Its Priority', *University of Chicago Law Review*, 40 (1973), pp. 534-55.
- Hill, T. E. Jr., 'The Stability Problem in *Political Liberalism*', *Pacific Philosophical Quarterly*, 75 (1994), pp.333-52.
- Horton, J., 'Rawls, Public Reason and the Limits of Liberal Justification', *Contemporary Political Theory*, 2 (2003), pp. 5-23.
- Ingram, A., 'Rawlsians, Pluralists, and Cosmopolitans', in *Philosophy and Pluralism*, edited by Archard, D. (Cambridge: Cambridge University Press, 1996)
- Jones, P., 'Two Conceptions of Liberalism, Two Conceptions of Justice', *British Journal of Political Science*, 25 (1995), pp.515-550.
- Klosko, G., 'Rawls's "Political" Philosophy and American Democracy', *American Political Science Review*, 87 (1993), pp. 348-59.
- Klosko, G., 'Rawls's argument from Political Stability', *Columbia Law Review*, 94 (1994), pp. 1882-97.
- Klosko, G., 'Political Constructivism in Rawls's *Political Liberalism*', *American Political Science Review*, 91 (1997), pp. 635-46.
- Kohlberg, L., 'The Claim to Moral Adequacy of a Highest Stage of Moral Judgement', *Journal of Philosophy*, 70 (1973), pp. 630-46.
- Krasnoff, L., 'Consensus, Stability, and Normativity in Rawls's Political Liberalism', *The Journal of Philosophy*, 95 (1998), pp. 269-92.
- Kukathas, C. and Pettit, P., *Rawls: A Theory of Justice and its Critics* (Cambridge: Polity Press, 1990).
- Larmore, C., *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987).
- Larmore, C., 'Pluralism and Reasonable Disagreement', *Social Philosophy and Policy*, 11 (1994), pp.61-79, reprinted in Larmore, C., *The Morals of Modernity* (Cambridge: Cambridge University Press, 1996)
- Larmore, C., 'The Moral Basis of Political Liberalism', *The Journal of Philosophy*, 96 (1999), pp.599-623.
- Larmore, C., 'Public Reason', in Freeman, S. (ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003).
- Lloyd, S.A., 'Relativizing Rawls', *Chicago-Kent Law Review*, 69 (1994), pp. 709-35.

Macedo, S., 'Liberal civic education and religious fundamentalism: the case of God v. John Rawls?', *Ethics*, 105 (1995), pp.248-64.

Mandle, J., 'The Reasonable in Justice as Fairness', *Canadian Journal of Philosophy*, 29 (1999), pp.75-107.

Marneffe, de P., 'Liberalism, Liberty and Neutrality', *Philosophy & Public Affairs*, 18 (1990), pp.253-74.

Martin, R., *Rawls and Rights* (Lawrence: University Press of Kansas, 1985).

Martin, R., 'Rawls's New Theory of Justice', *Chicago-Kent Law Review*, 69 (1994), pp. 737-61.

McClennen, E. F., 'Justice and the Problem of Stability', *Philosophy and Public Affairs*, 18 (1989), pp. 3-30.

Michelman, F.I., 'Rawls on Constitutionalism and Constitutional Law', in Freeman, S. (ed.), *The Cambridge Companion to Rawls* (Cambridge, Cambridge University Press, 2003).

Mitchell, L.E., 'Trust and the Overlapping Consensus', *Columbia Law Review*, 94 (1994), pp.1918-35.

Moller Okin, S., 'Reason and Feeling in Thinking about Justice', *Ethics*, 99 (1989), pp. 229-49.

Moller Okin, S., *Justice, Gender, and the Family* (New York: Basic Books, 1989).

Moller Okin, S., 'Political Liberalism, Justice, and Gender', *Ethics*, 105 (1994), pp. 23-43.

Mulhall, S. and Swift, A., *Liberals and Communitarians*, 2nd edn (Oxford:Blackwell, 1996)

Mulhall, S. and Swift, A., 'Rawls and Communitarianism', in Freeman, S.(ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003).

Nagel, T., *Equality and Partiality* (New York: Oxford University Press, 1991).

Pogge, T.W., *Realizing Rawls* (New York: Cornell University Press, 1989)

Quong, J., 'The Scope of Public Reason', *Political Studies*, 52 (2004), pp. 233-50.

Raz, J., 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy and Public Affairs*, 19 (1990), pp. 3-46.

Reidy, D.A., 'Rawls's Wide View of Public Reason: Not Wide Enough', *Res Publica*, 6 (2000), pp. 49-72.

Richards, D.A.J., 'Public Reason and Abolitionist Dissent', *Chicago-Kent Law Review*, 69 (1994), 787-842.

Scheffler, S., 'The Appeal of Political Liberalism', *Ethics*, 105 (1994), pp. 4-22.

Tronto, J., *Moral Boundaries. A Political Argument for an Ethic of Care* (New York: Routledge, 1993).

Van Parijs, P., 'Social Justice and Individual Ethics', *Ratio Juris*, 8 (1995), pp. 40-63.

Walker, B., 'John Rawls, Mikhail Bakhtin, And The Praxis of Toleration', *Political Theory*, 23 (1995), pp.101-27.

Waldron, J., 'Disagreements About Justice', *Pacific Philosophical Quarterly*, 75 (1994), pp. 372-87.

Waldron, J., 'Judicial Review and the Conditions of Democracy', *The Journal of Political Philosophy*, 6 (1998), pp. 335-55.

Wenar, L., '*Political Liberalism: An Internal Critique*', *Ethics*, 106 (1995), pp. 32-62.

Weithman, P. J., 'Liberalism and the Political Character of Political Philosophy', in *Philosophy of Rawls. A Collection of Essays*, volume 5 *Reasonable Pluralism*, ed P.J. Weithman (New York: Garland Publishing, 1999).

Wingebach, E., 'Unjust Context: The Priority of Stability in Rawls's Contextualized Theory of Justice', *American Journal of Political Science*, 43 (1999), pp.213-32.

Wolff, R. P., *Understanding Rawls. A Reconstruction and Critique of A Theory of Justice* (Princeton: Princeton University Press, 1977).