

JUSTICE AND THE DEMANDS OF REALISM

by

Daniel K. Munro

B.A. Political Science
University of Toronto, 1996

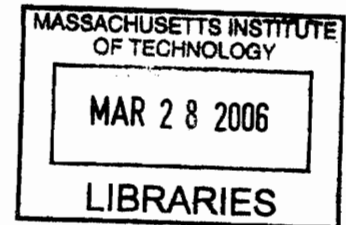
M.A. Political Science
University of Western Ontario, 1997

Submitted to the Department of Political Science
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Political Science

at the

Massachusetts Institute of Technology

February 2006



© Daniel K. Munro. All rights reserved.

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: _____
Department of Political Science
October 3, 2005

Certified by: _____
Joshua Cohen
Leon and Anne Goldberg Professor of Humanities
Professor of Philosophy and Political Science
Thesis Supervisor

Accepted by: _____
Roger D. Petersen
Associate Professor of Political Science
Chair, Graduate Program Committee

ARCHIVES

JUSTICE AND THE DEMANDS OF REALISM

by

Daniel K. Munro

Submitted to the Department of Political Science
on October 3, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Political Science

ABSTRACT

The dissertation examines how concerns about the demands of realism should be addressed in political theories of justice. It asks whether the demands of realism should affect the construction of principles of justice and, if so, how principles should respond to those demands. To address the problems posed by the demands of realism, the dissertation focuses on two specific realist concerns – namely, a concern about the motivational demands of justice and a concern about moral and religious pluralism – and asks what role, if any, these concerns should play in the articulation of principles of justice. Through a critical interrogation of the theories of Jean-Jacques Rousseau, John Rawls, and Jürgen Habermas, the author argues that a conception of justice should be attentive to citizens' motivational capacities and reasonable moral and religious worldviews but should not automatically accommodate their determinate preferences, opinions, and beliefs which may be objectionable. Endorsing an ideal of deliberative democracy and a conception of deliberative citizenship, the author argues that institutional arrangements which encourage democratic deliberation can help citizens to be more reflective about their determinate motives and beliefs and help them to acquire the desire and reason necessary to support just principles and institutions. At the same time, however, the author holds that appropriate institutions will be stable only when citizens acquire the necessary motivation and reason which leads to the defense of a dynamic model of justice, motivation, and reasonable pluralism in which just institutions and a just social ethos are regarded as mutually reinforcing. The author concludes that theorizing about justice should be limited not by what is given by the social and political status quo, but instead by the limits fixed by political hope.

Thesis Supervisor: Joshua Cohen
Title: Leon and Anne Goldberg Professor of Humanities
Professor of Philosophy and Political Science

For Heather

Acknowledgements

Working on a dissertation in which the word “motivation” appears on nearly every page can be enormously frustrating for an occasionally (often?) unmotivated graduate student. Luckily, I have been surrounded by wonderful people who at different times and in different ways have motivated and encouraged me to complete this work and to make it something that I am proud to call my own.

I begin by thanking my advisor, Joshua Cohen, who has been a great source of inspiration and support for the past seven years. Not only has he helped to make this a better work than it otherwise would have been, but his own writing and teaching have provided me with rigorous and inspiring standards which I can only wish to meet in my own writing and teaching. I am especially grateful to him for the example he has provided of a democratic political philosophy limited only by the idea of hope.

I would also like to thank the two other members of my committee. I thank Sarah Song whose critical remarks were always accompanied by words of (surely undeserved) praise and who generously blurred the boundaries between mentor, colleague, and friend. I also thank David Estlund whose enthusiasm for the project is deeply appreciated, whose contributions challenged me to make the argument more philosophically rigorous, and who, even on the worst of New England winter days, cheerfully made the journey from Providence to Cambridge to meet with us.

For valuable discussions and comments on earlier drafts, proposals, and half-baked ideas, I thank Karim Abdul-Matin, Marcos Ancelovici, Charles Beitz, Chedly Belkhodja, Michel Ben-Josef Hirsch, Suzanne Berger, Sharon Bilar, Gina Bishop, Cameron Bodnar, Corey Brettschneider, Kanchan Chandra, Francis Dupuis-Deri, Otis Gaddis, Richard Hansen, Joseph Heath, Sophie Hermann, Neil Hibbert, Charles Jones, Thomas McCarthy, Ibrahim Ouattara, Gabriella Rosen, Karen Rothkin, T.M. Scanlon, Neema Sofaer, Christine Telyan, Richard Vernon, Christine Wenc, Stuart White, and Amos Zehavi. I also thank G.A. Cohen for providing me with drafts of his unpublished work, for permission to quote from those drafts, and for discussions with him following his delivery of the Gilbane Lecture at Brown University in 2003.

A special thanks to my friend Loren King who has suffered through more than his share of my half-baked ideas and early drafts and who has helped to guide me through some of the more perplexing and frustrating features of academic life.

Without the help of Susan Twarog, Pam Clements, and Helen Ray I doubt that I would have been able to navigate the administrative, financial, and immigration complexities of my, at times, unconventional academic career. Thanks to each of them.

For their wonderful support and encouragement, I owe great thanks to my parents, Linda and Ken Munro, to my part-time political philosopher sister Cindy and her husband Dominic, and to other family-members and friends who have helped along the way. I was also encouraged and inspired by my late aunt, Donna Gaye McMorrin, and friend, James Dinneen. I miss you both.

Finally, and most importantly, I thank my wife, Heather. The personal sacrifices you’ve made in order to allow me to finish this project are more than I would dare to ask of anyone. That you made such sacrifices without my having to ask overwhelms me. For that, as well as your unfailing love, support, and power to turn my heart from despair to hope, I am forever grateful. I dedicate this work to you.

Table of Contents

Title.....	1
Abstract.....	3
Dedication.....	5
Acknowledgements.....	7
Table of Contents.....	9
1. Introduction: Justice and the Demands of Realism.....	11
Part I	
2. Rousseau: Monological Realistic Utopianism.....	55
3. Rawls: Hypothetical Constructivism.....	99
4. Habermas: Real Proceduralism.....	133
Part II	
5. Justice and Pluralism.....	179
6. Justice, Realism, and Deliberative Citizenship.....	215
Bibliography.....	253

1

Introduction: Justice and the Demands of Realism

Only in public life can we jointly, as a community, exercise the human capacity “to think what we are doing,” and take charge of the history in which we are all constantly engaged by drift and inadvertence....[T]he distinctive promise of political freedom remains the possibility of genuine collective action, an entire community consciously and jointly shaping its policy, its way of life.

Hanna Pitkin (1981: 344)

A conception of justice for a democratic society provides an account of the rights, liberties and resources required by citizens to pursue a conception of the good and to be full and equal participating members of democratic political life. It guides us on how institutions ought to be designed and arranged and how rights and resources should be distributed to ensure a fair system of social cooperation. Moreover, as an account of how institutions ought to be arranged and how resources ought to be distributed, it provides standards with which existing political arrangements and distributions can be assessed.

Even if we can agree on what a conception of justice is and what it is supposed to do in this general sense (and agreement on that is certainly hard to find), there is significant disagreement about how attentive to the social and political status quo any conception of justice should be when it is designed or articulated. In his discussion of global justice, for example, Charles Beitz considers the claim that morality and justice “cannot demand the impossible,” that the demands of justice should be limited to those which are motivationally possible or realistic (Beitz, 1999: 155). While Beitz agrees that justice should not demand the impossible in the sense of demanding that justice ignore or seek to overcome “impediments that are unalterable or unavoidable”, he maintains that “impediments to change that are

themselves capable of modification over time” should not be included in our notion of the impossible (1999: 156). We should respect the limits of the possible when theorizing about justice, Beitz holds, but we should not constrain ourselves by a social and political status quo which is “capable of modification over time.”

While Beitz points to an important distinction between demands of realism which are fixed and demands of realism which are capable of modification, it is not entirely clear how, or even if, that distinction should play a role in theorizing about the content of justice. Why might one think that the demands of realism – in either fixed or malleable form – should play a role in theorizing about justice at all? Perhaps theorists of justice should ignore both the fixed and malleable demands of realism. By contrast, if we decide that either the fixed or malleable demands of realism matter to theorizing about justice, in what way exactly should those demands affect the content of a conception of justice?

My aim in the dissertation is to examine how concerns about the ideal and the real should be addressed in political theories of justice. That is, I ask whether features of the social and political status quo – what we can call the demands of realism – should affect the construction of principles of justice and, if so, how principles should respond to those demands. To address the problems posed by the demands of realism in a manageable and rigorous manner, however, the dissertation focuses on two specific features of human beings and the world that might affect judgments about the realism of a conception of justice – namely, the motives and motivational capacities of human beings, on the one hand, and the pluralism of moral and religious doctrines in contemporary liberal democracies, on the other.

In the first case I ask what role, if any, concerns about motivation should play in the articulation of principles of justice and I treat this problem as a piece of the larger puzzle about the demands of realism. In particular, I ask whether there are any motives that should

be accommodated in a conception of justice and I discuss whether such accommodations are to be made for reasons of justice, reasons of practicality, or both. Additionally, I investigate the limits that might be imposed on the demands of justice once we consider agents' motivational limitations. In short, I ask whether the content of a conception of justice should be adjusted to account for what citizens are *already* motivated to support or what they are motivationally *capable* of supporting.

In the second case, I ask what role the existence of religious and moral pluralism should play in the articulation or construction of principles of justice and this too is treated as a piece of the larger puzzle about the demands of realism. With respect to pluralism, I ask first whether the simple existence of a diversity of moral and religious views – the fact of pluralism, as Rawls calls it – requires changes in a conception of justice. But I also ask, perhaps more significantly, whether the content of certain kinds of comprehensive doctrines – namely, *reasonable* comprehensive doctrines – should affect the content of principles of justice. That is, I ask not only whether the comprehensive views that currently exist require accommodation, but whether there are reasons to adjust a conception of justice in light of the range of reasonable comprehensive views that citizens could endorse.¹ Moreover, I inquire into the ways in which a defensible conception of justice might fit into the different worldviews of citizens and thereby increase support for the demands of justice when those demands are seen as demands which citizens have reason to support. Indeed, I ask whether the content of a conception of justice should be adjusted to account for what citizens believe they *already* have reason to support, or what they *could* have reason to support, in light of the moral or religious worldviews to which they are or could be committed.

¹ I thank Dave Estlund for repeatedly pressing me to recognize a distinction between the empirical fact of pluralism and the range of possibly reasonable comprehensive doctrines and for indicating

There are other features of the social and political status quo which we could have considered in thinking about the demands of realism. So, for example, one might consider how to deal with the form of the family in theorizing about justice. Should the family be taken as a sphere of private action beyond the scope of justice, should the relations between members of families be regulated by a conception of justice, or should there be some kind of middle ground? One might consider the challenges posed by an existing economic system or distribution of property. Should a conception of justice take the market system as a fixed constraint on principles or should that system itself be subject to normative evaluation? Should a conception of justice work with the existing distribution of property as a constraint or should it give no weight at all to existing distributions? Moving away from more formal institutions and arrangements, one might consider the complications of theorizing about justice in light of resource scarcity or features of one's geographic and demographic condition such as an aging population or, conversely, an overwhelmingly young population.

Thus, concerns about motives and comprehensive doctrines – the two features of the social and political status quo that I wrestle with in the dissertation – constitute just two possible demands of realism among many others. But thinking about these two features and whether and how they might be accommodated in a conception of justice will help us to reach more general conclusions about how a conception of justice should deal with the demands of realism. I have chosen these two concerns in particular because of their connection to issues and ideas about citizenship and about what human beings are like and what they are capable of. That is, motives and beliefs are features of the political landscape that are tied to human beings and for which human beings may bear some responsibility because they are thought, at least on some level, to be the products of human decision.

what difference the distinction could make for a conception of justice. I will address this more fully

Focusing on these demands of realism, then, requires us to navigate carefully through both the possibility and desirability of deliberative transformation of certain demands of realism in the way that more fixed demands of realism – such as natural scarcity or geographic and demographic realities – would not. To be sure, because I have chosen to focus on demands of realism that are not fixed in the same way that natural scarcity or geographic realities are, I have ensured that how we deal with the demands of realism is not solved merely by pointing to a distinction between what is fixed and what is malleable. Indeed, the depth of the problem is revealed more clearly by a focus on demands of realism whose fixity is itself in question. Thus, it may be the case that certain motives and beliefs should be accommodated in a conception of justice, rather than transformed, even though such motives and beliefs are *not* fixed and *could* be transformed. Or it may be that because motives and beliefs are amenable to transformation that they should have no bearing on the construction of a conception of justice.

In any case, the focus on motives and beliefs as part of the larger concern about the demands of realism ensures that the dissertation addresses a wider range of theoretical issues and possibilities. And, moreover, the focus on motives and beliefs allows us to consider the relationship between justice and citizenship – that is, it allows us to think through the ways in which justice depends upon ideas and assumptions about citizens' concerns and capacities and the ways in which conceptions of citizenship are shaped by the requirements of justice. Indeed, as I will argue, a conception of justice should confront the demands of realism with an ideal of deliberative citizenship in the background and a conception of justice should be designed in ways which might help to shape and reinforce the deliberative capacities and democratic commitments of citizens.

Thus, the dissertation is animated by an attempt to answer a set of questions with respect to the relationship between justice and the limits of the possible in general and between justice, motivation, and pluralism in particular: *Why* should concerns about motivation and doctrinal pluralism have a role in theorizing about justice at all? *How* are we to distinguish between motives or motivational limitations which should and those which should not be accommodated in a conception of justice? *How* are we to distinguish between comprehensive doctrines, or features of comprehensive doctrines, which should and those which should not be accommodated? And how should we respond to the problems of motivational possibility and doctrinal compatibility if we fail to accommodate citizens' existing motives and features of citizens' comprehensive doctrines in the conception of justice?

To answer these questions, I examine how concerns about motivation, pluralism, and realism more generally, are treated in the theories of Jean-Jacques Rousseau, John Rawls, and Jürgen Habermas. Each of these theorists takes the demands of realism seriously and tries to offer a conception of justice and political arrangements which are, to borrow a term from Rawls, "realistically utopian". A "realistically utopian" approach to political philosophy, Rawls writes, "extends what are ordinarily thought to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition" (1999b: 11). Each of these theorists attempts to reconcile utopian ideals with the demands of realism without simply resigning principles of justice to an objectionable status quo. We want principles of justice that will play a role in the political and social lives of real human beings, they hold, rather than utopian principles that have no connection to the motives, motivational capacities, and reasonable moral and religious doctrines of human beings.

But while these theorists think that principles of justice should have some practical worth, they share a commitment to the idea that principles should be concerned above all with how we *ought* to arrange our political institutions. Nevertheless, while each theorist provides some clues about how the tension between justice and the demands of realism should be addressed, I argue that, for different reasons, each view ultimately fails to provide an adequate solution to that problem. As we will see, Rousseau's *monological* variant of realistic utopianism, which holds that normative principles should be those we discover through a process of inner reflection, doesn't take seriously the normative implications of the range of reasonable pluralism. Rawls's *hypothetical constructivist* variant, which holds that principles should be the outcome of a process of reflection on relevant reasons, fails to offer a sufficiently democratic account of legitimacy. And Habermas's *real proceduralist* variant of realistic utopianism, which regards justified principles as those which emerge from an actual process of democratic deliberation, fails to offer a compelling account of how the institutional configuration which realizes the procedure will deliver the necessary motivational possibility and doctrinal compatibility. In the course of critical examination of these theories, then, I develop my own approach to navigating the demands of realism that attempts to overcome the weaknesses inherent in the monological, hypothetical constructivist, and real proceduralist variants of realistic utopianism.

In the remainder of this introductory chapter, I intend to do the following: First, I will offer a preliminary account of what we mean when we speak of motives and I will explain in greater detail what I take to be the main problem with respect to justice and motivation (section I). In section II, I will introduce and clarify the ideas of pluralism and reasonable pluralism and describe the challenge posed to justice by the fact and range of reasonable pluralism. Having explained the problems of motivation and pluralism, I will indicate in

section III what other theorists believe to be at stake and how the concerns about motivation, pluralism, and realism more generally should be addressed, if at all. We will explore how thinkers who lean towards a utopian view believe that facts about human beings and the world should have no effect on the content of justice while thinkers who lean towards a realist view think that substantial accommodations of those features of the status quo must be made if a conception of justice is to be practical. In section IV, I will offer a preliminary sketch of my solution to the justice and realism tension. And finally, in section V, I will offer an outline of the chapters and further describe the ideas of the three theorists whose work will be at the centre of critical attention.

I. Motives and the Problem of Motivation

In a recent article on issues of motivation in neo-Kantian theories of justice, Vittorio Bufacchi offers a useful definition of the term ‘motive’:

[a] motive is what induces a person to act. The Latin root of the word is *motives*, meaning ‘to move’, hence the idea of movement, or initiating action. A motive gives a person a reason to perform an action. It follows that a moral motivation may be defined as what induces a person to act morally, or what reason a person has to act morally (Bufacchi, 2005: 26).

While Bufacchi’s definition captures the core idea of what a motive is, we should be careful not to confuse the idea that a motive “gives a person a reason to perform an action” with T.M. Scanlon’s view that having a motive to act (morally) is a matter of having or recognizing a reason to act that could be justified to others (Scanlon, 1998: 147-158). If that is what Bufacchi means by motive – and his discussion of Scanlon’s work suggests that might be his meaning – then we should not endorse the definition as a whole. The Scanlonian aspect of the definition imports a normative test rather than merely describing or defining the term on its own.

Surely people can have motives which are not and perhaps could not be framed in terms of reasons that are justifiable to others. A person may be motivated by malevolence, cruelty, or a desire for relative advantage over others, none of which would likely be acceptable to others. Even less malign motives, such as self-preservation and restricted compassion, might not be reasonable in the Scanlonian sense. So while Bufacchi's definition of a motive as "what induces a person to act" is acceptable, we should not confuse the *definition* of motive – i.e., that which gives a person "a reason to perform an action" – with a *normative* test for reasonable motives – i.e., that having a motive is a matter of having a reason that could be justified to others. If we want to work through the problem of motivation understood as a part of the larger problem about justice and the demands of realism, then we should adopt a definition of motive which does not narrow the scope of the realism problem by importing a normative test for reasonable motives at the outset. For our purposes then, a motive will be defined as whatever moves an individual to act.

There are various things that might move a person to act and therefore, a variety of things that might count as motives. Agents may be moved to act by their perceived interests, preferences, emotions, or desires. Moreover, these different sources of motivation may be united at times and thus together move an agent to perform a particular action whereas at other times, perhaps most of the time, motives may conflict with each other and compete to be the motive which ultimately determines action. With that in mind, I will use the term motive in an expansive, rather than restricted, sense. A restricted use of the term motive would hold that only those things which have in fact induced an agent to act are motives, so that we would count as motives only those things which induce, or have induced, determinate acts. A more expansive use of the term, and the one I will use here, holds that we include as motives not simply the last motive in deliberation which in fact causes a

person to act – what we might call the operational motive – but rather any motive which is, or what the agent considers to be, a *possible* inducement to action.

The more expansive use of motive has an advantage over the restricted use insofar as it allows us to endorse the possibility of motivational plurality – the possibility that agents have sets of potential motives, rather than single motives which determine all of their behaviour, and that the operational motive is simply one among many. Moreover, the expansive use of the term helps us to make sense of a state of internal motivational *unity*, on the one hand, and a state of internal motivational *conflict* on the other – that is, the possibility that an agents' potential motives agree on a determinant action and the possibility that an agent's operational motive may nevertheless be in conflict with the agent's other potential motives.

The problem, then, is that there may not be a fit between what individuals are motivated to do and what a conception of justice requires them to do. To be sure, what motivates agents' behaviour is sometimes regarded as a constraint on what can be realistically achieved or demanded by a conception of justice. If human beings are naturally and primarily motivated by narrow self-interest, for instance, a principle of justice which calls for significant redistribution of economic resources will be difficult or impossible to implement. Absent external incentives for compliance or sanctions for noncompliance, people may simply choose to ignore the demands of such principles even in cases where they think that the principles themselves are good. Some theorists think that it is rational for them to do so, even if they regard such noncompliance as unjust. The prudent thing to do, one might say, is not always the right thing to do.

Thus, certain motives appear to be in tension with the requirements of candidate principles or conceptions of justice.² An individual will often be faced with choices between doing what she is already motivated to do and doing what she ought to do according to some conception of justice. Just as existing motives appear as constraints on the demands of justice, the demands of justice appear to individuals as constraints on their choices and behaviour. From the perspectives of both justice and ordinary citizens, then, what is desirable may not be attainable given the constraints imposed by motives, in the former case, and by the demands of justice in the latter case.

To illustrate, imagine that a conception of justice requires an equal distribution of resources among all citizens but that, at present, resources are unequally distributed. The required redistribution in that case would conflict with some citizens' desires to give special consideration to their own children and to use their larger present holdings to provide them with health and educational advantages, for example. The equal distribution requirement, then, confronts those citizens as an external constraint on what they are permitted (and motivated) to do for their children. Resources that they would like to use for their children's benefit instead are transferred to anonymous others in order to satisfy the equal distribution requirement of the conception of justice. In short, there are cases in which the demands of justice will conflict with the preferences and desires which motivate citizens' behaviour.

There may be very good reasons for placing constraints on the permissible actions of citizens or for refusing to make concessions in the principles of justice themselves simply to accommodate certain existing motives. But even if there are good reasons for those

² I have used the term "candidate" principles or conceptions of justice here to avoid the mistaken impression that the dissertation is concerned with how the demands of realism interact with *already* settled or known principles of justice. The dissertation explores, among other things, whether or not motives and beliefs should influence the content of principles of justice *prior* to the recognition or

constraints, citizens might nevertheless still regard and feel them *as constraints* rather than as reasonable limitations on their permissible actions (or they might regard them as both). Sometimes citizens will acknowledge the demands of justice and change their behaviour accordingly (perhaps because they see the demands as reasonable or perhaps because they regard the punishments for flouting the demands as something to avoid). At other times, citizens will plump for their own concern over the demands of justice. In either case, many citizens will regard the demands of justice as demands which are in competition with their existing motives. Meanwhile, from the perspective of justice, citizens' motives will be regarded as the constraints. A conception of justice might set out requirements for a fair and just society, for example, but find that those requirements can't be met because of the practical constraints imposed by citizens' existing motives and capacities.

Political philosophers generally agree, then, that motivation affects the possibility of realizing or maintaining a conception of justice. Whether or not a given conception of justice is realistic depends, among other things, on the capacity and motivation of human beings to adopt and to live by that conception. This problem of motivation is sometimes framed in terms of a "rift" (Cronin & De Greiff, 1998: xii) or a "gap" (Williams, 1985: 52) between the demands imposed by a conception of justice, on the one hand, and the desires and motivational capacities of human beings on the other. As Cronin and De Greiff observe, "[t]he unrestricted universality of moral principles, their highly abstract, cognitive claims to validity, and the unconditional character of the duties they impose create a rift between moral judgment and reasoning, on the one hand, and motivation, on the other" (1998: xii). When a conception of justice makes demands on human beings that conflict

adoption such principles and not what we should do when motives and beliefs conflict with already settled principles.

with the determinate content of their existing motives, the conception of justice is for that reason less realistic than a conception which does not make similarly strong demands.

However, while we might agree that the concern for motivation presents justice with a certain demand of realism, there is substantial disagreement about how to deal with such a demand, if it should be dealt with at all. Should theorists try to make room for agents' motives and hope that doing so will ensure greater stability for the principles of justice? Or should theorists ignore motives in order to insulate principles of justice from objectionable features of the social and political status quo? Or, otherwise still, should theorists try to distinguish between motives which should and those which should not be accommodated in a conception of justice? Before we attempt to answer those questions, consider another example of the way in which the demands of realism complicate theorizing about justice.

II. Pluralism and Justice

While motives and motivational capacities account for one manifestation of the way in which features of the world confront a conception of justice with certain demands of realism, the existence of a diversity of moral and religious doctrines in liberal democracies presents conceptions of justice with another example of a demand of realism. Whereas the problem with motives is that agents have desires or preferences which may conflict with the demands of justice, the problem with moral and religious pluralism is that people may have moral or religious *reasons* for acting contrary to the demands of a candidate conception of justice.

So what exactly is a moral, religious, or comprehensive doctrine, and what do we mean by the terms fact of pluralism, fact of reasonable pluralism, and range of reasonable pluralism? Before we discuss the nature of the pluralism challenge, then, we should clarify

what we mean by pluralism, reasonable pluralism, and comprehensive doctrines or worldviews.

A comprehensive doctrine, as Rawls defines it, is a moral or religious conception that “includes conceptions of what is of value in human life, and ideals of personal character, as well as ideals of friendship and of familial and associational relationships, and much else that is to inform our conduct, and in the limit to our life as a whole” (1996: 13). Whereas an individual may develop or endorse a conception of the good which indicates “what is of value in human life” and “what is regarded as a fully worthwhile life”, a comprehensive doctrine is the “religious, philosophical, or moral” worldview which offers an account of how “the various ends and aims are [to be] ordered and understood” and according to which such conceptions are interpreted (2001: 19). In contemporary political societies we find a variety of such comprehensive doctrines or worldviews which not only offer different accounts of what is a valuable or a worthwhile life, but often accounts which conflict with those offered by other moral and religious worldviews. That is, contemporary political societies are characterized by a fact of pluralism – the fact that there are a variety of different, incompatible, and perhaps conflicting moral and religious comprehensive doctrines to which citizens adhere or aspire.

So, for example, one’s moral or religious worldview might regard homosexuality as sinful and thus might require one to support and to advocate policies which would ensure that homosexual behaviour is restricted, perhaps severely punished, and that individuals who identify as homosexuals should be regarded as morally repugnant. By contrast, a candidate conception of justice might require that all citizens be treated as equals and that behaviour which does not directly harm others should be tolerated. In this case, then, what justice demands of all citizens conflicts with what some citizens believe they have reason to do and

to think and, moreover, what sort of sanctions and punishments it is permissible (indeed required) to impose on others. In this case, we might hold that the adherents of the religious view in question are unreasonable in the sense that they refuse to acknowledge other citizens as free and equal citizens with concerns and commitments of their own. But merely regarding those citizens as unreasonable does not alter the fact that those citizens do not regard themselves as having a good reason to adhere to the demands of justice in question.

The fact and range of *reasonable* pluralism describe more challenging conditions. The fact of reasonable pluralism entails, first, that a subset of those different moral and religious views in society are compatible with fair political arrangements. That is, the fact of reasonable pluralism entails that the subset of reasonable views are held by people who are prepared to agree to fair terms of social cooperation, even with others who hold different worldviews, and that their being reasonable in that sense is compatible with the moral or religious doctrine to which they subscribe. As Rawls puts it,

The diversity of reasonable comprehensive religious, philosophical, and moral doctrines found in modern democratic societies is not a mere historical condition that may soon pass away; it is a permanent feature of the public culture of democracy. Under the political and social conditions secured by basic rights and liberties of free institutions, a diversity of conflicting and irreconcilable – and what's more, reasonable – comprehensive doctrines will come about and persist if such diversity does not obtain already (1996: 36).

Thus, while conceptions of justice may require certain commitments or behaviour from citizens, the fact that citizens have moral and religious worldviews which may not map directly onto a shared conception of justice means that the demands of justice may conflict with the demands of those moral and religious views. Citizens may have reasons – not unreasonable moral and religious reasons – to act in ways contrary to candidate conceptions of justice. Indeed, their beliefs about the world and what is of value may require them to think and behave in one way whereas the demands of justice may require them to behave in

an entirely different way. Moreover, not only must a conception of justice wrestle with the already existing reasonable comprehensive doctrines – the *fact* of reasonable pluralism – it must also wrestle with the *range* of possibly reasonable comprehensive doctrines. That is, there may be a range of reasonable moral and religious worldviews that a conception of justice may be required to accommodate even if those reasonable worldviews are not currently embraced by any existing individuals. The idea of a *range* of reasonable pluralism, then, suggests that a conception of justice may have to be constructed in ways which anticipate what I refer to as normatively salient or moral facts even if those normatively salient or moral facts are not currently instantiated in the world as we now find it.

The fact of pluralism and the fact and range of reasonable pluralism in democratic societies, then, confront a conception of justice with what appears to be a dichotomous choice: One can choose to ignore the pluralism of ethical, religious, and philosophical views and insist on independent standards of justice untainted by the demands of pluralism. Or one can adjust the conception to accommodate the fact of pluralism or the range of reasonable pluralism so that principles of justice can play a role in practical political life and thereby avoid the charge of naïve utopianism. Theorists who choose the former option can be criticized for ignoring a permanent and perhaps attractive feature of democratic societies. Indeed, some of the moral and religious worldviews might be endorsed by reasonable citizens who are sincerely committed to regarding other citizens as free and equal citizens and to locating fair principles of justice for a plural society. Theorists who choose to adjust principles to accommodate pluralism, on the other hand, might succeed in solving the problem of doctrinal compliance but risk making concessions to unattractive or objectionable features of certain worldviews that we find in contemporary plural democracies. The challenge is to find some alternative to the dichotomy. How can we

accommodate pluralism without giving up on the normative and critical value of a shared conception of justice? Or should we aim to accommodate pluralism or reasonable pluralism at all? What exactly is necessary to reconcile a conception of justice to pluralism and, moreover, to what extent (if any) must a conception accommodate the determinate beliefs of citizens?

III. Justice and the Demands of Realism: Utopian and Realist Views

Recognition of the tension between justice and facts about human beings and the world has led some theorists to attempt to bracket what they regard as concerns about mere practicality or realism while articulating or constructing principles of justice to ensure that attention to those concerns does not affect the content of justice in objectionable ways. John Stuart Mill puts the point with particular force when, in defending his utilitarian conception of justice against those who charge that the standard is “too high for humanity” and therefore disconnected from reality, he replies that “this is to mistake the very meaning of a standard of morals, and confound the rule of action with the motive of it.” “The business of ethics,” Mill advises us, is “to tell us what are our duties, or by what test we may know them,” not to accommodate the standard of justice to meet concerns of practical possibility (Mill, 1991: 149).

Others, like Hobbes (1991) and Hume (1983, 1985), wary of the utopian tendency to ignore human beings’ desires and capacities and other features of the social and political world, respond to the tension between justice and the demands of realism by calling for more, rather than less, attention to those demands.³ As Hume writes, “custom is the great

³ In his critique of Rousseau, J.L. Talmon criticizes the utopian tendency to “train men to ‘bear with docility the yoke of public happiness’, in fact to create a new type of man, a purely political creature, without any particular private or social loyalties, any partial interests” (1952: 42). Similarly, Karl

guide of human life” (1983: 46) and “men must endeavour to palliate what they cannot cure” (1985: 38). Recognizing that principles of justice will achieve stability only when human beings have the necessary desire and reason to uphold and to live by the principles of justice, realist theorists contend that the best way to ensure a fit between principles, desires, and reasons is to design principles in the full light of what human beings are already motivated to support and what human beings already regard themselves as having reason to support. To be sure, some theorists hold that external moral norms which do not already have a place in the motivational sets of moral agents cannot provide the necessary motivation. Instead, these theorists look to existing self-interested motives and beliefs as the raw material for constructing a stable conception of justice.⁴

David Miller suggests that in the face of the gap between what some political philosophers think justice requires and what ordinary people *think* justice requires political philosophers should pay more attention than they do now to “what the people think” (1999: 61). This is especially the case, Miller writes, for “the normative theorist who is guided by something akin to the Rawlsian ideas of reflective equilibrium and public justifiability” (1999: 59). As he puts it, “in setting out a theory of justice, [a Rawlsian theorist] needs evidence about what people do in fact regard as fair and unfair in different social settings” (1999: 50). The idea seems to be that if there is too wide a gap between what normative theorists think justice requires and what ordinary people believe justice requires, then ordinary people will lack the reason and motivation necessary to uphold and live according to the more demanding conception of justice.

Popper calls for “piecemeal” rather than “utopian engineering” because the former is more democratic and more attentive to circumstances and interests. With piecemeal engineering, he writes, “[t]here will be a possibility of reaching a reasonable compromise and therefore of achieving the improvement by democratic methods” (1966: 158-9).

But Miller makes a move that other theorists would regard with skepticism: Miller suggests that the fact that people hold or share certain beliefs about justice gives us reason to acknowledge and to perhaps accommodate those beliefs in normative principles. Or, at least, that these shared beliefs should be taken as evidence of the kind of “considered judgments” which play a role in reaching reflective equilibrium about a conception of justice. It is not clear at the outset, however, that the fact that certain beliefs about justice are *shared* gives us a strong reason to accommodate those beliefs in a conception of justice. Recall Beitz’s distinction between impediments to change that are themselves *capable of modification over time*; and impediments that are *unalterable* and *unavoidable* (1999: 155-6). The fact that certain beliefs and motives exist and that they might be shared by many people would not give us, on Beitz’s view, reason to adjust principles of justice to accommodate those beliefs and motives in part because those beliefs and motives might be capable of modification.

There are other theorists who would go further than Beitz in this regard and argue that it is unacceptable to distort principles of justice to accommodate agents’ motives and beliefs, whatever those might be. G.A. Cohen (2000, 2003) pursues this position when he argues that to offer an account of justice which makes concessions to features of human behaviour which are the products of unequal relations of power, defective institutions, or unreflective conceptions of human nature – rather than assessing and trying to alter behaviour and institutions – is to give up on the essential normative and critical functions of political philosophy. Indeed, he asks why we should “respond to the discrepancy between the putative and attractive conception of justice and people’s moral capacities by assigning a defect to that conception of justice...rather than to people’s moral capacities?” (2003: 122).

⁴ If I understand the project correctly, this is the impetus for Gauthier’s (1986) attempt to provide a rational basis for morality.

If the most defensible conception of justice is unrealistic given agents' motives and beliefs, then perhaps all we should do is condemn the human race for its inherent injustice.

Assurance, Compliance, and Justice

The positions defended by Mill (1991), Beitz (1999), and Cohen (2000, 2003) appear to others as too uncompromising – and not simply because they are unrealistic, but because they fail to see how some demands of realism in fact operate as considerations of justice. Consider the so-called assurance problem which is often invoked in discussions about justice and legitimate obligations. Some theorists hold that if we are to expect people to adhere to the demands of a conception of justice which requires some amount of self-sacrifice or restraint, then we must be able to assure those people that such restraint “will be matched by similar restraint on the part of others” (Miller, 1999: 19). People have an obligation to be just, this view holds, only if they can be given assurance that their sacrifice or restraint will not be used by others as an opportunity and occasion for exploitation. If reciprocity is a constitutive feature of justice, then there could be no obligation to meet the demands of justice if we lack assurance that others will meet those obligations as well.

Thus, in defending Rawlsian global justice against critics like Beitz, Joseph Heath argues that “[g]iven the central role that cooperation and reciprocity play in Rawls’s system, the absence of the rule of law at an international level is not merely a ‘practical’ difficulty. It plays a central role in determining what individuals can reasonably expect of one another under such circumstances” (Heath, 2005: 7-8). Institutions, such as the rule of law, are the mechanisms through which the assurance problem can be solved. If those institutions are not in place, then agents have no obligation to meet the requirements of justice because those requirements would put them at risk of exploitation by other agents whose self-

interested motives cannot be checked. In effect, this means that the requirements of justice are shaped or reshaped in light of the existing institutional possibilities for overcoming the problem of motivational compliance. If existing institutional arrangements fail to protect agents whose motivations are compatible with the rigorous demands of justice from agents whose motivations are comparatively less benign, then the well-intentioned agents are absolved of some, perhaps all, of the demands of justice.

While we might agree that the limits of institutional *possibilities* could make a difference in the content of principles of justice, Heath seems to perpetuate the realist conflation of the limits of the probable with the limits of the possible. While it is true to say that how far we can and how far we should try to realize the demands of justice under given conditions is determined in large measure by the existing institutional and motivational terrain, it seems too much to say that absent the appropriate institutional and motivational conditions the content of justice itself changes.

In some ways, Heath's realist argument makes the same assumptions about the nature of justice that G.A. Cohen's more utopian argument makes. Both seem to think that justice is concerned with individual behaviour – what agents are obligated, or not obligated, to do in their personal behaviour – in contrast to the Rawlsian view that justice is concerned, at least in the first instance, with the basic structure of society. Heath seems to hold that I have no reason to transfer a certain amount of my current holdings to another, as a conception of justice may require, if institutional arrangements which would protect me from exploitation by the immoral are not in place. In that case, he thinks, justice does not demand supererogatory acts. But the content of justice would not change under those conditions if we thought that justice was a matter not simply of individual decisions and behaviour concerning resource distributions, but that justice required individuals to work towards

establishing and stabilizing the institutional conditions necessary for the other demands of justice to be realized.⁵ In that sense, we return to Beitz's distinction between impediments that are amenable to modification and impediments that are unalterable. To solve the problems of motivational and doctrinal compliance, then, Beitz holds that we should not just take account of the existing institutional, motivational, and doctrinal reality and shape justice to fit that reality. Instead, we should consider the institutional, motivational, and doctrinal possibilities and take those *possibilities* as providing the parameters within which we theorize about justice.

Yet, while Beitz's approach might respond adequately to the realist position which draws the limits of the possible too narrowly, it still faces challenges from the utopians who seem to think that little or no attention should be paid to practical issues when theorizing about the fundamental principles of justice. Thus, even if we think that it would be desirable for a conception of justice to avoid accommodating motives and beliefs, or to avoid accommodating *objectionable* motives and beliefs, we might ask whether it would also be desirable to avoid consideration of practical concerns altogether. G.A. Cohen seems to think that theorizing about justice should avoid such questions altogether and that, if they do enter the picture, they should enter only at the level of application, not at the level of justification or articulation of principles of justice. But we should ask whether that conclusion is defensible. That is, even if we decide to insulate conceptions of justice from the effects of malign or objectionable motives and beliefs, should we also insulate conceptions of justice from concerns about motivational possibility – about whether, even under ideal conditions, agents have the motivational *capacity* to support the demands of a

⁵ This puts me at odds with the Rawlsian idea that we take the institutional conditions – the basic structure – as settled in a certain sense. See Heath (2005, 2005b) on this.

conception of justice? And should we insulate conceptions of justice from concerns about the limits of moral and religious transformation?

As Bernard Williams argues, it is not enough to articulate a conception of justice using something like a Rawlsian veil of ignorance and asking what rational agents, ignorant of their particular preferences, aims, and identities, would choose. “The question ‘what do rational agents need’ is not enough to lead those agents into morality” (1985: 64). On Williams’ view we need to ensure that citizens are, or could be, motivated to be moral or just and that they have, and recognize that they have, reason to be just. Thus, we also need to consider the problem of how we might generate or locate agents’ motivation to be just or, as Rawls calls it, the “sense of justice” (1999: 41, 125; 1996: 19). Even if agents lack at present the motivation to be just, must a defensible theory of justice offer an account of how those agents *could* be motivated? Must we also consider how citizens who endorse or could endorse different, but perhaps reasonable, moral and religious views, might locate reasons to be just that are compatible with their moral and religious views? Or, alternately, must we demonstrate how it would be possible and legitimate to transform those moral and religious views in more justice-supporting ways?

Theorists like Cohen and Mill, as we have seen, seem less troubled by those questions than are theorists like Hobbes, Hume, Williams, and Elster. For Williams and Elster, one must show how citizens can come to endorse the principles even when their present motives and beliefs conflict with those principles. Framing the problem as one of social order, Elster asks “What is it that glues societies together and prevents them from disintegrating into chaos and war?” (1989: 1). Williams, concentrating more than Elster on the normative significance of the problem, asks how obligations of justice can “stick” to political agents and operate as motivating aims of their action (1981: 122). So while Williams and Elster

might agree with Beitz that we should not regard “impediments that are...capable of modification over time” as determining the limits of the possible, they nevertheless seem to think that an account of how those impediments can be overcome is required.

Sanctions for Noncompliance

Before I offer a sketch of the solution that I defend in the dissertation, I want to consider briefly another proposal which is often invoked as a way to deal with the problem of how to motivate agents to support and abide by the demands of justice. We can call this the sanctions-for-noncompliance approach. Assuming that we have already settled on a particular conception of justice, this approach holds that perhaps the best way to ensure that all citizens comply with the demands of that conception – and thus the best way to quell fears of exploitation by free-riders – would be to impose penalties or sanctions on those who fail to comply with the demands. And, to make the system run smoothly, we should ensure that the sanctions for noncompliance outweigh any expected benefits from noncompliance.

While sanctions for noncompliance certainly have an important role to play in assuring the stability of a conception of justice, they fail to ensure that individuals will be motivated to uphold the conception of justice when that is necessary. As I will argue throughout the dissertation, and at length in Chapters 4 and 6, the sanctions-for-noncompliance approach fails to take account of the fact that the laws and sanction mechanisms must themselves be created and maintained by citizens in view of what justice demands. If the sanctions-for-noncompliance approach is to enforce *justice* and not some arbitrary and perhaps indefensible set of laws, then those citizens who create and maintain the system of sanctions must themselves have a sense of justice and be motivated to design and maintain the system with the conception of justice as a guide. Thus, while the sanctions-for-noncompliance approach

has an important role, it is insufficient on its own as a solution to the problem of motivation because such a system can only enforce justice if it has been designed to do so.

A conception of justice which lacks a plausible account of motivational possibility, as we will see, may fail to live up to its own hope for justice. Though external rewards and punishments might be necessary in the practical application of principles of justice where the motivation isn't already developed, a conception of justice which relies exclusively on such external factors faces great difficulties in proving that it is not only legitimate and just, but that it is stable in the long term as well. Solving the problem of motivation, then, requires an account of how the motivation necessary to support a given conception of justice can either be located or generated in an *internal* way rather than simply imposed through external sanctions or constraints.

Moreover, a conception of justice which fails to locate an overlapping consensus of reasonable comprehensive doctrines on a shared conception of justice will fail also to achieve stability and, for that matter, justice. Indeed, citizens who are committed to moral or religious worldviews may be prepared to suffer significant sanctions for noncompliance if they think that adherence to their moral or religious commitments is nonnegotiable. While imposing sanctions raises the costs of violating the demands of justice for those citizens, high legal or economic costs for violation may not weigh in the deliberations of some citizens at all. Avoiding high legal or economic costs, that is, will not be regarded by them as a reason to do or not to do something. When the facts of pluralism and reasonable pluralism are considered in theorizing about justice, then, we must wrestle with the *reasons* that citizens have for compliance or noncompliance and attempt to locate some justice-supporting reasons that they could endorse.

A Formal Statement of the Realist-Utopian Disagreement

Before proceeding to the sketch of my proposed solution to the justice-realism tension, consider the following set-up of the nature of the disagreement between the realist and utopian views.

It might be said that a conception of justice has or should have two basic features:

[1] A conception of justice must offer *critical standards*, and

[2] A conception of justice must be *practical*.

In the first place, a conception of justice provides, or should provide, critical standards [1]. That is, it must set out an account of those norms or values with which we ought to comply if we or our institutions are to be considered *just*. As I noted in the opening paragraph to this chapter, a conception of justice for a democratic society provides an account of the rights, liberties and resources required by citizens to pursue a conception of the good and to be full and equal participating members of democratic political life and, as a result, provides standards with which existing political arrangements and distributions can be assessed.

In the second place, a conception of justice should be practical [2]. That is, a conception of justice should be, or must be, action-guiding in the sense of offering an account of justice which agents are *able* to follow or realize.

But in thinking about the practicality requirement we should recognize that what people do depends on what they *desire* and on what they regard themselves as having *reason* to do. That is, an agent's compliance with a conception of justice depends on her having the *desire* or *motivation* to comply and *reason* to comply. Or, at the very least, that her compliance depends on her having non-justice related desires and reasons which have the effect of

producing justice-complaint behaviour whether she aims at justice or not. With the addition of this observation, our set-up of the problem looks like this:

[1] A conception of justice must offer *critical standards*.

[2] A conception of justice must be *practical*.

[3] As a matter of fact, what people *do* depends on the *desires* and *reasons* they have.

While an ideal typical *realist* would accept [1], the realist thinks that [2] must be given greater weight in theorizing about justice once [3] is acknowledged. Indeed, the acknowledgment of [3] is what prompts the realist to give priority to [2] over [1]. That is, the realist thinks that a conception of justice should be practical and accepts that what people do depends on what they desire and what they have reason to do. In that case, for a conception of justice to be practical, the realist thinks that existing desires and reasons will have to be accommodated in some way. For the realist, then, the *practical* aim [2] trumps the *critical* aim [1] when the alleged fact that action depends on desires and reasons [3] is acknowledged.

The ideal typical *utopian*, by contrast, accepts [1] and [3], but rejects [2] altogether. The utopian agrees with the realist that what people do depends on what they have desire and reason to do, but holds that a conception of justice must offer critical standards rather than *mere* action-guiding or practical principles. For the utopian, then, the *critical* aim [1] trumps the *practical* aim [2] *despite* the fact that action depends on desires and reasons [3]. It seems that we are compelled to choose between the realists' giving priority to [2] over [1] or the utopians' endorsement of [1] and outright rejection of [2].

In the dissertation I explore the possibility that such a dichotomous choice between [1] and [2] might not be required in light of [3] and, therefore, that it might be possible to have a

conception of justice which is both practical and critical while acknowledging that [3], or something like it, is true. Thus, while I accept the basic intuitions behind [1], [2], and [3], I aim to show that they are in fact compatible once we properly understand what each of them really entails. To anticipate, my argument holds that when we acquire a better understanding of the ideas expressed in [1], [2], and [3], the apparent need to pursue either the ideal typical realist or ideal typical utopian track is not necessary at all. Whereas the common understanding of the practicality condition [2] assumes that we must rely on what people are already *able* to do, I argue that satisfying the practicality condition (in those cases where it should be satisfied) requires only that people are *capable* of meeting the demands of a conception of justice. And whereas condition [3] holds that what people do depends on their existing desires and reasons, I argue that what people are *able* to do depends not on their *existing* desires and reasons, but on those desires and reasons that they are *capable* of having. Which means that while condition [3] may be true, it does not give us the whole story; a further clarification allows us to weaken the apparent force of [3] as stated above. Thus we get the following set of compatible conditions:

[1] A conception of justice should offer *critical standards*.

[2]' A conception of justice should be practical in the sense of offering action-guiding principles with which agents are *capable* of complying.

[3] What people do depends on the desires and reasons they have, but

[4] What people are *able* to do depends on the desires and reasons that they are *capable* of having.

In short, once we get a better understanding of what condition [2] entails and add clarification [4], we see that the conditions are not incompatible as both the realist and utopian hold. Thus, we are forced to endorse neither the realist nor the utopian approach to justice and the demands of realism. But whether those revisions are compelling and what they really amount to requires further elaboration.

IV. Towards a Realistically Utopian Solution

It seems, then, that there are two levels to the problems of motivation and pluralism – one level is concerned with what we should do about existing motives and beliefs while the other is concerned with how (and for that matter whether) we should demonstrate motivational possibility, on the one hand, and doctrinal compatibility, on the other. The solution that I defend in the dissertation is directed at both levels. To solve the first-level problem – the problem of existing motives and beliefs – we need an argument about which motives and beliefs should or must be accommodated by a conception of justice and those which should not and, additionally, an argument about why such accommodations are necessary at all. To solve the second-level problem we need a plausible account of how the motivation necessary to support the conception of justice can be generated and, second, a plausible account of how citizens might locate support for a conception of justice from within their reasonable moral and religious worldviews.

Existing Motives and Beliefs

One of the central claims that I defend in the dissertation is that there are certain motives and beliefs that should be regarded not as mere practical constraints on the feasibility or stability of justice, but instead as constitutive elements of justice itself. Certain motives and

beliefs must be accommodated as a matter of justice and not simply as a matter of prudence or stability. That view seems to put me at odds with the utopian view which holds that concerns about the social and political status quo are entirely practical rather than normative. To alter the content of justice to meet the demands of realism, they hold, is to resign ourselves to objectionable features of our social and political status quo. By contrast, I argue that some motives and some beliefs are *normatively* salient. Citizens who are motivated by a concern for their liberty, for example, are motivated by an attractive idea which should not be dismissed so hastily as a mere practical matter. Similarly, the concerns of citizens who endorse reasonable ethical or religious worldviews should not be dismissed by theorists of justice. The range of reasonable pluralism, I will argue, must be accommodated not for reasons of prudence, but for reasons of justice. In short, I argue that once certain freedoms are recognized as unobjectionable or even valuable, a conception of justice for a democratic society must regard those freedoms and some of their consequences as candidates for accommodation.⁶

Defending the idea that certain motives and beliefs must be accommodated as a matter of justice rather than mere practicality will require a method for distinguishing between motives and beliefs with normative salience and those without normative salience. I argue that such distinctions should be made from the point of view of a society of free, equal, and reasonable citizens capable of critical reflection on their motives, beliefs, and capacities. In short, the distinction should be made from the point of view of an ideal of deliberative

⁶ David Estlund has suggested to me the idea that if we think that there are certain features of human beings and the world that need to be accommodated by a conception of justice because they are *normatively salient*, then in fact we are not “accommodating” those features or “capitulating” to certain demands of realism; rather, we are doing precisely what justice requires. I continue to use the term accommodation not because I disagree with Estlund; rather, I use the term because it seems a less cumbersome way to refer to the idea that certain motives and beliefs must count as legitimate considerations in the construction or articulation of a defensible conception of justice.

democracy and an associated conception of deliberative citizenship. Those motives and beliefs which should be accommodated, then, are those which reasonable citizens would say are necessary for the protection of their freedom and their desire to ensure fair conditions for the pursuit of their reasonable conceptions of the good. When candidate principles of justice conflict with what is necessary for the protection of citizens' freedom and the fair conditions for the pursuit of reasonable conceptions of the good, then those candidate principles must be adjusted, not for reasons of mere feasibility or stability, but in the interests of justice itself.

But if the motives and beliefs that we accommodate are accommodated because they are normatively salient, then it might be said that the utopian position prevails. That is, so far we have not made any accommodations for reasons of practicality and thus haven't moved away from the utopian position which holds that practical concerns are irrelevant to theorizing about justice. Where we do move away from the utopian position and begin to flesh out a distinct *realistically* utopian view is with our consideration of the second-level problems of motivational possibility and doctrinal compatibility.

Motivational Possibility, Doctrinal Compatibility, and Deliberative Citizenship

Even after the normatively salient motives and beliefs have been accommodated, there are still certain features of human beings and the world which act as constraints on the possible demands of justice. That is, there are logically possible candidates for demands of justice which would be practically impossible for citizens to fulfill because they lack not simply the motivation at present, but the capacity to do so. A conception of justice, I will argue, must reconcile itself not only to the normatively salient motives of human beings but to the limits of their capacities as well. Indeed, because the solution to the problem of existing motives

and beliefs defended here fails to accommodate *all* antecedent motives and beliefs, an adequate approach to concerns about realism will also have to offer a solution to the problem of motivational possibility. Though I will not try to show how a particular conception of justice should locate that motivational possibility, I will argue that any compelling conception of justice will have to show how it can generate or draw on *internal* sources of motivation and/or on the moral and religious doctrines which citizens endorse.

To put it another way, a defensible conception of justice must offer a solution to the problem of motivational possibility which does not rely on external constraints or coercive means to assure its stability. That means that a conception of justice must be able to show how citizens living under institutions designed according to that conception could develop the sense of justice necessary to support those institutions and that conception.

To solve the problem of motivational possibility, then, I shall do three things: First I will argue that the kind of reasonable reflection on antecedent motives, which is demanded by the ideal of deliberative democracy, is consistent with a realistic conception of deliberative citizenship. I try to show that human beings are capable not simply of strategic reasoning on determinate aims and interests, but that human agents have a capacity for practical, reflective reasoning about the necessity and value of those aims and interests which is facilitated and encouraged by deliberative democratic arrangements.

Second, solving the problem of motivational possibility requires institutions which are designed not only in accordance with the demands of the favoured conception of justice but which can also encourage the exercise of deliberative citizenship and generate both a greater sense of legitimacy and motivational support for the conception of justice. Institutional arrangements which encourage democratic deliberation, I argue, can help citizens to be more reflective about their determinate preferences, opinions, and beliefs and help them to acquire

the “sense of justice” necessary to support a conception of justice. At the same time, however, I maintain that appropriate institutions will be adopted and stable only when citizens already have the necessary motivation to adopt and support those institutions which leads me to endorse a dynamic model of justice, motivation, and reasonable pluralism in which just institutions and a just social ethos are mutually reinforcing.

Third, and finally, I argue that a conception of justice can and should draw support from citizens’ attachment to reasonable ethical or religious comprehensive doctrines but in a way which avoids unreflective and unnecessary accommodations of the determinate content of any particular view. When citizens are able to regard a conception of justice as compatible, perhaps even required by, their particular moral and religious worldviews, that conception of justice, I argue, will be more stable because regarded as legitimate. Thus, I articulate and defend a more democratic version of Rawls’s idea of an overlapping consensus which seeks to reconcile political agents to the fact of reasonable pluralism and to generate legitimacy for a freestanding political conception of justice. While the appeal for support from reasonable comprehensive doctrines may not improve the motivational realism of a conception of justice, it would give citizens reasons to endorse a conception of justice that are themselves more immediately accessible to those citizens.

Justice, Realism, and Legitimacy

If those arguments are compelling then we will have made significant progress towards offering a compelling approach to the concerns about justice and realism. But an important question will remain: Why should addressing concerns about realism matter at all when theorizing about justice? The solution to the first-level problem – the problem of existing motives and beliefs – has a built-in response: We accommodate those motives and beliefs

which are normatively salient rather than merely practically salient. But the solutions to the problems of motivational possibility and doctrinal compatibility lack that kind of built-in response. Indeed, from the utopian perspective the problems of possibility and compatibility appear as problems of mere application or stability rather than normatively salient problems. Worrying about these problems seems to confuse fundamental principles of justice with the application of those principles.

Throughout the dissertation I shall try to make the case that solving the problems of motivational possibility and doctrinal compatibility, which appear to be mere stability problems, are in fact requirements of justice. A conception of justice that lacks a plausible, even if improbable, account of how it could preserve itself over time would fail to gain the endorsement of democratic citizens concerned not with their private preferences, interests, and beliefs but with the common good of a democratic society.

The idea is this: A conception of justice, on my view, must be a conception to which citizens agree or could agree. But even if we can locate that agreement, once the conception of justice is in place, it is likely that some citizens will violate or be tempted to violate the demands of justice from time to time. In that case, a just society is justified in using or threatening to use coercion to ensure compliance with the fair terms of social cooperation which the violators or potential violators themselves have, or could have, endorsed as reasonable. But surely no one could rationally or reasonably agree to principles of justice which are either motivationally impossible or with which reasonable comprehensive doctrines are incompatible because doing so would be to agree to be subjected to inevitable and futile coercion. That is, if the demands of justice are beyond the motivational capacities of human beings or require commitments and behaviour which reasonable citizens attached to reasonable comprehensive doctrines could not fulfill, then citizens would *inevitably* face

coercive action. And, moreover, because changes in behaviour and attitude are either motivationally or doctrinally impossible, the coercion would serve no practical purpose, it would be *futile*. My claim, then, is that because no rational or reasonable citizen would agree to unavoidable and pointless coercive action, only a conception of justice which offered solutions to the problems of motivational possibility and doctrinal compatibility could achieve legitimacy.

In the end, motivational possibility and doctrinal compatibility are not simply concerns about the realism or stability of a conception of justice, but concerns about the very *justness* of the conception of justice.

Justice, Realism, and Reasonable Hope

I asserted that solutions to the problems of motivational possibility and doctrinal compatibility must be provided as a matter of justice but I also asserted that the account of how citizens could be motivated to uphold a defensible conception of justice simply needs to be an account of possibility rather than one of probability. Yet the standard of possibility, rather than probability, might strike realist theorists as inadequate. Against that realist concern, I will make the case throughout the dissertation that our theorizing about justice should not be limited by what it is reasonable to *expect* in human behaviour and institutional design. Theorizing about justice should not be limited by the parameters of pessimistic or optimistic expectation. Rather, theorizing about justice should be guided, and limited, only by what it might be reasonable to hope for in political life.⁷

⁷ My interest in what I call the politics of hope and despair has a long personal history. It began with my first reading of Camus's essay, "The Myth of Sisyphus" (1991) some fifteen years ago. In that essay Camus instructs us that we must somehow imagine the tragic character Sisyphus happy. Additionally, the role that might be played by hope in political theory has been a background concern while reading Rousseau, Kant, and Rawls, among others. More recently, I was drawn into thinking

Thus, the distinction that should guide political theory is not a distinction between political optimism and pessimism, but instead a distinction between political hope and despair. Political optimism is the view that the societies in which we live will become more just as time passes; it is an expectation that citizens, institutions, and society as a whole will become more rather than less just over time. Political pessimism, by contrast, is the view that the societies in which we live will not become more just over time and that they may even become less so. But we should limit theorizing about justice not in the light what we *expect* or don't expect to occur. Rather, we should limit theory only by the parameters of political hope. Political hope entails the conviction that a better world is possible, even if unlikely. Though some may come to regard the demanding conception of deliberative citizenship defended here as evidence of a naïve political optimism, we should instead regard it as part of an attempt to construct an attractive conception of justice rooted in the idea that justice is within the reach of human beings. It is rooted in the hope that it is possible to sever the Thrasymachian link between power and the content of justice. As Rawls puts it, “political philosophy is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition” (1999b: 11). In short, a realistically utopian approach to justice limits its hopes only by what is practically possible, not by what is politically probable nor by what is given by the political status quo. Indeed, if theorizing about justice takes its cues from an

more seriously about hope and despair after reading Jose Saramago's devastating novel *Blindness* in which Saramago somehow manages to shine a light of hope into an otherwise depressing portrait of the human condition at the end of the twentieth century (Saramago, 1997). While these and other works, along with events in the world both tragic and pathetic, have made thoughts about the meaning and worth of hope an ongoing theme in my own internal monologue, my decision to work through these thoughts more systematically and publicly was inspired and encouraged by Joshua Cohen. For the example he has provided through his own work and the explicit encouragement he has given to me in this regard, I owe him an enormous debt.

unjust political status quo, and regards that status quo as unalterable, then it is hard to see how theorizing about justice could produce anything more than a political theory of despair.

Thus, a realistically utopian approach to principles of justice does not assume that all features of the world, including existing human desires and beliefs, are fixed. Some of those features can be changed even if the outer limits of change are themselves fixed. So what we should take as the limits of the possible, then, are not the existing motives and beliefs of human beings; rather the limits of motivational possibility and doctrinal compatibility should be fixed only by what we might reasonably hope for.

But what would it be *reasonable* to hope for? Why would it be reasonable to hope for certain changes and unreasonable to hope for others? Answering those questions in this introductory chapter would take us too far into the substance of the dissertation at this early stage. Nevertheless, asking the question now indicates that what we conclude about the role that the demands of realism, especially motivational and doctrinal concerns, should have in theorizing about justice, will be guided by what could be regarded as reasonable. As the appeal to an ideal of deliberative democracy and a conception of deliberative citizenship will make clear, the content of the reasonable is given by what free and equal citizens could justify to each other under ideal deliberative conditions. And that method of gleaning the content of the reasonable should ensure that when we consider the relationship between justice and reality we do not give in to the realists' pessimism, the utopians' naïve optimism nor, especially, to the politics of despair.

V. Chapter Outline

The first three chapters following this introduction examine how concerns about the demands of realism are addressed in the political theories of Rousseau, Rawls, and

Habermas. In these three chapters the focus is primarily on the first demand of realism – the problem of motivation – and while pluralism does receive some attention, it is addressed more fully in chapters 5 and 6. As I noted earlier, each theorist takes the practical aim seriously and tries to offer a conception of justice or political arrangements which would reconcile utopian ideals and the demands of realism without simply resigning principles of justice to an objectionable status quo. And while each theorist provides some clues about how the tension between justice and the demands of realism should be addressed, I argue that, for different reasons, each view ultimately fails to provide an adequate approach to solving that tension.

In Chapter 2, I consider Rousseau’s attempt to solve the problem of motivation with a combination of transformative and accommodationist strategies. Rousseau defends a political theory which accommodates citizens’ fundamental interests in freedom and security but also holds that interests and motives which conflict with the fundamental interests are legitimate objects of political transformation. In offering his solution to the problem of motivation Rousseau provides us with three critical ideas: the idea of a shared point of view which helps to distinguish between interests and motives which should and those which should not be accommodated by a conception of justice; a conception of the person understood as free and equal; and a view about the way institutions and the social ethos interact to produce the sense of justice necessary to solve the problem of motivational possibility.

However, while Rousseau offers these insights about how to resolve the justice and motivation tension, there are two weaknesses in his theory that prevent us from endorsing his view altogether: First, his move to identify the “fundamental interests” of human beings prior to democratic deliberation risks making citizens the objects rather than subjects of

political discourse and, consequently, unable to make contact with the sources of motivation on which his theory depends. Second, Rousseau draws the limits of practical possibility too broadly when he fails to acknowledge the limits imposed by the range of reasonable pluralism and the need to reconcile a conception of justice to that range. In short, while Rousseau offers some direction, the monological variant of realistic utopianism that he defends – according to which we discover appropriate principles through inner reflection – fails to adequately address concerns about justice and the demands of realism.

Like Rousseau, Rawls regards the tension between justice and motivation as significant and in need of a solution. Whether a conception of justice can achieve motivational stability is treated not as a mere practical issue but as a normatively salient concern. But like Rousseau, Rawls does not want to offer a conception of justice which amounts to a mere rational bargain among political actors fully aware of their determinate preferences, interests, and motives. Instead, he wants to articulate principles of justice which are insulated from agents' objectionable preferences, interests, and motives – that is, principles which maintain a critical distance from the social and political status quo and its resource and power inequalities.

Thus, in Chapter 3, I take up Rawls's *hypothetical constructivist* variant of the realistically utopian project according to which principles are presented as the outcome of a hypothetical procedure that identifies and relies on relevant reason. In that chapter I ask whether Rawls's attention to the "strains of commitment" leads him to make accommodations to unattractive motives and I argue that, in light of Rawls's later revisions of the role of the primary goods and the conception of the person, attention to the strains of commitment does not force concessions to objectionable motives. In the course of examining Rawls's view, we confront more fully the idea that motivation matters not simply for reasons of feasibility and stability,

but for reasons of justice itself. Indeed, Rawls defends a realistically utopian conception of justice which in part ties the justification of principles to their legitimacy and, in so doing, indicates that motivational concerns are relevant to the content of justice.

However, two concerns about Rawls's *hypothetical constructivist* approach are left unaddressed in Chapter 3: In the first place, it is unclear why legitimacy and agreement need to play such fundamental roles in the construction of a conception of justice. As G.A. Cohen (2003) holds, it is not clear what, if anything, legitimacy and agreement add to the justness of a conception of justice. Perhaps concerns about motivation can be handled without the idea of agreement. The other outstanding issue has to do with how Rawls's approach fares in light of the diversity of moral and religious worldviews which characterizes contemporary democratic societies – the other feature of our social and political condition with which the dissertation is concerned. In the search for motivational possibility under conditions of pluralism, Rawls's political theory might require further revision. A final assessment of Rawls's approach to justice and the demands of realism, then, is postponed until these outstanding issues are addressed.

Chapter 4 begins with Habermas's attempt to address the first unaddressed issue from the previous chapter and it reconsiders Rawls's political theory in light of the more democratic approach to the problem of motivation that Habermas defends. Habermas champions a discursive conception of democracy which tries to embrace Rousseau's conviction, on the one hand, that "human beings act as free subjects only insofar as they obey just those laws they give themselves" (Habermas, 1996: 445-6) and Rawls's claim, on the other hand, that principles of justice should not be tainted by objectionable interests nor influenced by power inequalities. At the same time, however, Habermas rejects Rawls's attempt to insulate deliberation about principles of justice from the determinate preferences,

interests, and beliefs of citizens in an original position. Instead, he defends a principle of discourse ethics which gives citizens a central role in discourses about those norms which will apply to institutions and actions while it simultaneously filters out objectionable motives and interests via an impartial moral point of view implicit in the pragmatic presuppositions of argumentation. Thus, Habermas offers what we might refer to as a *real proceduralist* approach, as opposed to the Rawlsian *hypothetical constructivist* approach, to dealing with concerns about justice and realism

While Habermas's attempt to democratize the procedure for discovering appropriate norms is an important development and one which a Rawlsian approach should consider, Habermas's *real proceduralist* variant of the realistically utopian approach ultimately fails to provide an attractive approach to the justice and realism tension. In his effort to "unburden" citizens of the cognitive, motivational, and organizational demands of complex societies by shifting from inclusive discourses to a more anonymous flow of reasons (Habermas, 1996: 114), Habermas abandons the mechanism by which transformations of citizens' pre-political preferences and beliefs can be achieved and which can encourage the development of the "postconventional stage of moral consciousness" (Habermas, 1990: 109) or sense of justice. I argue that a concern for citizens' capacity to support valid principles of justice is unavoidable once we recognize that the very institutions which Habermas thinks can sidestep the realism problem ultimately depend on citizens' sense of justice and their designing and reforming institutions in light of that sense of justice.

With the critical examinations of Rousseau's *monologicalism*, Rawls's *hypothetical constructivism*, and Habermas's *real proceduralism* essentially complete, the dissertation moves to a more systematic consideration of the pluralism concern in Chapter 5 and towards more constructive aims in Chapter 6. In Chapter 5, I consider more directly the problems raised

by the pluralism of moral and religious views in democratic societies. I elaborate and defend a more democratic version of Rawls's overlapping consensus of reasonable comprehensive doctrines and, in doing so, defend the idea that beliefs and reasons matter not simply for the stability of a conception of justice, but for the very legitimacy and normative attractiveness of a conception of justice. Additionally, I begin to sketch a conception of deliberative citizenship which indicates how the revised idea of an overlapping consensus is neither naïvely utopian nor objectionably practical. That conception of the deliberative citizen, I will argue, helps us to see how we can practically accommodate pluralism in democratic societies without losing hope that a just society of free and equal citizens is possible.

Finally, in Chapter 6, I further elaborate and defend my own view about how the demands of realism should be addressed in theorizing about justice. The problems of motivation and pluralism should be addressed, first, by accommodating some, but not all, antecedent motives and beliefs and that decisions about which should and which should not be accommodated are to be made from the point of view of an ideal of deliberative democracy and a conception of deliberative citizenship. Thus, the solution to the problem requires, second, a richer description and defense of the idea of deliberative citizenship sketched at the end of Chapter 5 and an argument about how that conception helps to solve the problems of motivation and pluralism. Additionally, in Chapter 6, I explain how the interaction between institutions and citizens' ethos demonstrates why motivational and doctrinal concerns are matters of justice and not merely matters of stability and I point to a solution to the problem of motivational possibility which draws on the idea of deliberative citizenship.

Rousseau: Monological Realistic Utopianism

The limits of the possible in moral matters are less narrow than we think. It is our weaknesses, our vices, our prejudices that shrink them.

- Jean-Jacques Rousseau (1978: 3.12.2)

Let us therefore begin by setting all the facts aside, for they do not affect the question.

- Jean-Jacques Rousseau (1964: 103)⁸

Rousseau's attempt to resolve the tension between justice and the demands of realism moves in a realistically utopian direction. His theory tries to capture the utopian intuition that a conception of justice should avoid accommodations to objectionable motives while simultaneously responding to the realist challenge to articulate a conception which would be motivationally stable. The "fundamental problem" that Rousseau sets for his political philosophy, then, emerges from the thought that principles of political right must be responsive to some demands of realism and unresponsive to others. In his words, the problem is to "find a form of association that defends and protects the person and goods of each associate with all the common force, and by means of which each one, uniting with all, nevertheless obeys only himself and remains as free as before" (*SC*: 1.6.4). Rousseau's contractual solution to that problem places human beings and their most basic interests, as well as an idea of legitimacy, at the centre of political theory. Indeed, not only does Rousseau think that a conception of justice must ensure respect for citizens' fundamental

⁸ Rousseau's works are cited in this chapter as follows: *SC* 1.6.4.: *On the Social Contract*, ed. Roger D. Masters, trans. Judith R. Masters (New York: St. Martin's Press, 1978), Book 1, chapter 6, paragraph 4; *LD*: *Letter to d'Alembert on the Theatre*, trans. Allan Bloom (Ithaca: Cornell University Press, 1968); *D1*: *First Discourse*, and *D2*: *Second Discourse*, ed. R. Masters, trans. R. Masters and J. Masters (New York: St. Martin's Press, 1964); *E*: *Emile*, trans. Barbara Foxley (London: Everyman, 1993); *P*: *The Government of Poland*, trans. Willmore Kendall (Indianapolis: Hackett Publishing Company, 1985).

interests, it must also offer an account of motivational possibility. That is, if agents are not already motivated by their objective fundamental interests, as Rousseau thinks they should, then the account of motivational possibility he offers helps us to see how it would be possible for citizens to be so motivated.

But Rousseau is not a brute realist who thinks that existing preferences, interests, and motives must be accommodated uncritically by a conception of justice. While he wants political theory to have a practical role and thinks that principles must be both constrained and legitimized by the limits of motivational possibility, he does not want to make concessions to objectionable or unattractive motives. In the *Second Discourse* Rousseau urges us to “begin by setting all the facts aside, for they do not affect the question” (D2: 103). The *Social Contract*, moreover, holds that it is both possible and desirable to transform the social ethos of citizens to be more supportive of just social arrangements. Yet, the idea that we should set facts aside and work to transform the social ethos leads some critics to complain that Rousseau’s political theory is objectionably utopian and that it threatens, rather than protects, freedom. Indeed, it would appear that Rousseau’s transformative ambitions conflict with his aim to look for a legitimate and reliable social order “taking men as they are.” The tension between the projects of accommodation and transformation presses us to ask just what are the motives that, in Rousseau’s view, require transformation and what are the motives or interests which a conception of justice must respect?

Rousseau offers what we can call a *monological* variant of the realistically utopian approach to justice and realism. On that kind of view, principles of justice are to be discovered through a process of inner reflection. Thus, Rousseau relies on a moral point of view, which he calls the “inner voice” of reasonable human beings, to distinguish between human beings’ real and artificial interests. The fundamental human interests in freedom and

security identified by the moral point of view not only require protection in political arrangements but give those political arrangements their central focus. Interests and motives which conflict with the fundamental aims of security and freedom are legitimate objects of political transformation, on Rousseau's view, and those transformations are achieved through socialization under just institutions. In effect, Rousseau defends a political theory which realizes and protects citizens' fundamental interests in freedom and security and provides an account of how that conception of politics could be motivationally stable. In doing so he provides us with three ideas that help to resolve the tension between justice and motivation: the idea of a shared point of view which helps to distinguish between motives which should and those which should not be accommodated by a conception of justice; a conception of the person understood as free and equal; and a view about the way institutions and the social ethos interact to produce the sense of justice necessary to solve the problem of motivational possibility.

However, while Rousseau offers these insights about how to resolve the justice and motivation tension, there are two weaknesses of his *monological* approach that prevent us from endorsing his view altogether: First, Rousseau's move to identify the "fundamental interests" of human beings through a process of inner reflection and thus prior to democratic deliberation risks making citizens the objects rather than subjects of political discourse and, consequently, unable to make contact with the sources of motivation on which Rousseau's theory depends. Second, Rousseau fails to acknowledge the limits of practical possibility imposed by the range of reasonable pluralism and the need to reconcile a conception of justice to that range.⁹ While reasonable citizens could come to endorse

⁹ The phrase "*range* of reasonable pluralism" was suggested to me by David Estlund. I use it rather than the term "*fact* of reasonable pluralism" used by Cohen (1993) and Rawls (1996) because the former allows us to think about not just the existing moral and religious worldviews and how they

Rousseau's political arrangements, the civil religion which accompanies those arrangements and which acts as an additional source of motivation is incompatible with the pluralism that would emerge from the political institutions. Where that pluralism is "reasonable" – an idea I will explain below – Rousseau faces not only a problem of motivational stability, but a problem of justice as well. In effect, while Rousseau's monological approach to the justice and realism tension offers useful insights with respect to motives and motivational possibility, the reliance on inner reflection makes it difficult for agents to recognize the range of reasonable pluralism which constitutes another important demand of realism to which a conception of justice must attend.

I will begin with an examination of Rousseau's projects of transformation and accommodation (section I). Next, I will consider the moral point of view which guides those projects (section II) and the solution Rousseau offers for the problem of motivational possibility (section III). In section IV, I present the first of two objections – namely, that in the absence of more deliberative and participatory arrangements Rousseau neglects an important source of motivational possibility. In section V, I present what I think is a more fundamental objection to Rousseau's attempt to reconcile justice and the demands of realism – namely, that adopting a *monological* approach to resolving the tension, and thus locating the moral point of view in the "inner voice" of human beings, fails to take account of the range of reasonable pluralism.

might affect the content of justice, but also the range of possibly reasonable comprehensive doctrines that might require accommodation even though they are not instantiated in the world as we find it now.

I. Transformation and Accommodation

Although Rousseau thinks that interests and motives are relevant to political theory – that political theory should have a practical role and that principles must be both constrained and legitimized by fundamental human interests – he does not want to make concessions to objectionable or unattractive interests and motives.¹⁰ Resigning ourselves to accommodating all existing interests and motives would be to make justice nothing more than a description of the social and political status quo with all its inequalities and restrictions on freedom. By contrast, Rousseau wants to offer an emancipatory political theory, a political theory that would instruct human beings on how to free themselves from enslavement to vanity and how to overcome the inequalities in political and economic power which contribute to their misery. And that emancipation requires a significant transformation of the social and political ethos. Indeed, in the *Social Contract*, Rousseau holds that it is both possible and desirable to transform the social ethos of citizens to be more supportive of just social arrangements.¹¹ Whereas natural human beings are essentially good but unable to reap the benefits of social cooperation and whereas social human beings reap some of the benefits of social cooperation but exhibit certain social vices, Rousseau aims to set up institutions which would transform human beings into virtuous citizens – that is, citizens who could benefit from social cooperation while acting in just or moral ways.

But transformation represents only one element, albeit an important element, of Rousseau's political theory. Rousseau is just as concerned that the principles of justice accommodate interests in security and freedom that are the most basic and most important

¹⁰ I will discuss the relationship between interests and motives in Rousseau's theory later in this chapter (Section III). For the time being, I will consider how Rousseau tries to accommodate and/or transform interests and motives understood as similar demands of realism.

¹¹ Rousseau argues, for example, that there should be a Legislator who "is capable of changing human nature...of transforming each individual...of altering men's constitution" (*SC* 2.7.3).

interests which he thinks ultimately motivate behaviour and whose accommodation is necessary if we are to have principles of *justice* at all. Nevertheless, just as the transformative aim is guided and limited by certain facts about human beings, the task of reconciliation is guided and limited by a distinction between kinds of interests and motives. In the *Second Discourse* Rousseau criticizes Hobbes's understanding of what constitute fixed, and what constitute contingent, features of human beings and maintains that some of those contingent features are objectionable. Rousseau's realistically utopian political theory, then, emerges from the dialectic between the aims of accommodation and transformation and is guided by a moral point of view. Before inquiring into the content of that moral point of view, we should consider more fully Rousseau's projects of transformation and accommodation.

1. *Transformation*

Rousseau's *Social Contract* presents an ambitious plan to transform social human beings into virtuous citizens.¹² As Habermas puts it, "Rousseau imagines the constitution of popular sovereignty through the social contract as a kind of existential act of sociation through which isolated and success-oriented individuals *transform* themselves into citizens oriented to the common good of an ethical community" (Habermas, 1998: 102). It is this plan which raises the suspicions of critics who see Rousseau's political theory as objectionably utopian and totalitarian. To be sure, their fears are fanned by Rousseau's statement that "the less the individual wills relate to the general will, that is to say customary conduct to the laws, the more repressive force has to be increased" (*SC*: 3.1.13).

But Rousseau thinks that a transformation of the social ethos is necessary because human beings, in their social condition, have developed a number of objectionable motives

or vices which compete with the common good. He says that “men are wicked; a sad and constant experience makes proof unnecessary” (*D2*: 208n2). That the status quo social ethos is *undesirable* – i.e., characterized by a widespread desire for relative social standing and by a concern for private preference rather than general interest – gives Rousseau a reason to advocate transformation. That the social ethos is *artifactual*¹³ – i.e., the product of social artifice rather than an essential feature of the deep structure of human nature and therefore malleable – gives Rousseau hope that transformation might succeed.¹⁴ He aims, then, to discover a legitimate social order that can produce a “remarkable change in man, by substituting justice for instinct in his behaviour and giving his actions the morality they previously lacked” (*SC*: 1.8.1). Rousseau does not think that the task is to peel back the artificial layers and return human beings to the idealized state of nature; rather the task is to take people as they are – naturally good but socially corrupted – and turn them into virtuous citizens.

But why does Rousseau think that a just social ethos, or a virtuous citizenry, is necessary? Why not think instead that political institutions should be adopted which separate powers and establish a system of checks and balances so that the conflicting self-interests can be directed to produce good outcomes?¹⁵ To see why not, consider the fourth

¹² Indeed, Rousseau sees as one of his primary aims the examination of “the act by which a people becomes a people. For this act...is the true basis of society” (*SC*: 1.5.2).

¹³ I take this term from Cohen and Rogers (1995).

¹⁴ But not infinitely malleable. The fundamental human interests in self-preservation and freedom are not malleable because they are part of the deep structure of human nature.

¹⁵ Machiavelli (1979) holds that human nature is both selfish and immutable and, as such, cannot be relied upon for the stability of a republic. The best we can do, he argues, is to create a system of divided powers with checks and balances and arrange the system so that people’s selfishness, when checked and balanced by the selfishness of others, will produce good outcomes. While less pessimistic about human nature, Hamilton, Madison & Jay (1982) agree that people can not be relied upon to always act in virtuous ways and so they also defend a system of checks, balances, and divided powers to stabilize the American republic.

of Rousseau's types of laws – what he calls “mores, customs, and...opinions” (*SC*: 2.12.5).

He writes that this type of law

is not engraved on marble or bronze, but in the hearts of citizens; which is the true constitution of the State; which gains fresh force each day; which, when other laws age or die out, revives or replaces them, preserves a people in the spirit of its institution, and imperceptibly substitutes the force of habit for that of authority (*SC*: 2.12.5).

The idea seems to be that the social ethos of the citizenry supports the laws and institutions and provides a kind of standard for judging when they need revision and what revisions should be made. Indeed, Rousseau seems to think that a virtuous social ethos is necessary for the stability of just arrangements and for their effective operation. In this respect he anticipates the conviction of the late American judge Learned Hand who held that “liberty lies in the hearts of men and women; when it dies there, no constitution, no law, no court can save it; no constitution, no law, no court can even do much to help it. While it lies there it needs no constitution, no law, no court to save it” (1952: 190). Both men think that primary support for freedom comes not from institutions, but from citizens themselves. When institutions decay, or when structural changes make them ineffective, the only sure guide to reforming institutions rests in the social ethos of the citizenry.¹⁶

Yet Rousseau's reasons for aiming at a transformation of the social ethos are not merely those of stability and effectiveness. He believes that a society cannot be just unless its citizens are also just because his ideal polis embodies an idea of the common good whose realization depends on the democratic will of the sovereign citizens. Because it is democratic – that is, because it relies on citizens themselves to say what is and what is not in their

¹⁶ Consider the words of Pitkin (1981): “Only in public life can we jointly, as a community, exercise the human capacity ‘to think what we are doing,’ and take charge of the history in which we are all constantly engaged by drift and inadvertence....[T]he distinctive promise of political freedom remains the possibility of genuine collective action, an entire community consciously and jointly

shared, general interest – citizens must be made aware of what are, and how they can properly pursue, their fundamental interests. Citizens must ask themselves not simply what would be best for the satisfaction of their private interest or what they would prefer, but instead what the general will requires. If “one always wants what is good for oneself, but one does not always see it” then a transformation of the social ethos can help citizens to see just what is in the common interest (*SC*: 2.3.1). In particular, the transformation of the social ethos would help citizens to see that “authentic” acts of the general will aim to protect their fundamental and shared interests in security and freedom and that they aim to do so equally.¹⁷ And seeing that would help citizens vote as Rousseau thinks they should. It would help citizens to understand that “[w]hen a law is proposed...what they are being asked is not precisely whether they approve or reject the proposal, but whether it does or does not conform with the general will that is theirs” (*SC*: 4.2.8). Rousseau’s transformation of the social ethos would help to ensure that “the characteristics of the general will are...in the majority” (*SC*: 4.2.9).¹⁸

To achieve the desired transformation of “wickedness” into virtuous citizenship, then, Rousseau suggests that our hopes rest on the possibility of a Legislator writing good laws which would effect the virtuous transformation of the social ethos. This mechanism aims to change those preferences and beliefs which motivate behaviour that is objectionable from the point of view of human security and freedom, in ways that would help citizens protect

shaping its policy, its way of life.” This idea will receive significant elaboration and defense in Chapters 4 and 6.

¹⁷ “[E]very authentic act of the general will...obligates or favors all citizens equally” (*SC*: 2.4.8).

¹⁸ Rousseau also thinks that justice is not achieved unless it is felt in the human heart. In this he anticipates the Kantian idea that what matters is not the consequence of an action but the purity of the motive of the agent who performs the action. In *Emile* Rousseau writes “justice and kindness are no mere abstract terms, no mere moral conceptions framed by the understanding, but *true affections of the heart* enlightened by reason (*E*: 235-6).

their fundamental interests consistent with the protection of the fundamental interests of others. Rousseau writes that

[t]he general will is always right, but the judgment that guides it is not always enlightened....Private individuals see the good they reject; the public wants the good it does not see. All are equally in need of guides. The former must be obliged to make their wills conform to their reason. The latter must be taught to know what it wants....From this arises the necessity for a legislator (*SC* 2.6.10).

The role and powers of the Legislator seem to raise a dark cloud over Rousseau's otherwise democratic project. He says that the Legislator should be "capable of changing human nature...of transforming each individual...of altering men's constitution" and, for those tasks, he is given the power to draft the laws (*SC* 2.7.3). Yet, while the legislator *drafts* the laws, he has no *legislative right* – the people retain the power to accept or to reject a law that is proposed by him. Even so, the legislator must have recourse to some power in order to get the people to agree to those laws which are in their best interest – that is, laws which transform them in positive ways. The problem here is that the legislator can use neither force (because that would conflict with men's freedom to obey only those laws which they give themselves) nor reason (because people do not yet have the virtuous "social spirit" which would help them to recognize and accept good laws). As Rousseau puts it, "the social spirit which should be the result of the institution, would have to preside over the founding of the institution itself; and men would have to be prior to laws what they ought to become by means of laws" (*SC*: 2.7.9). So the legislator "must necessarily have recourse to another order of authority, which can win over without violence and persuade without convincing" (*SC*: 2.7.9). That other "authority" is religion and I will discuss that in section IV.

2. *Accommodation*

Rousseau's view of what is required in order for there to be a just political order is insulated from citizens' existing interests and the structures of power. It sets out a procedural and substantive test for principles and laws if they are to be regarded as legitimate which requires that citizens act from a knowledge of and commitment to a "general will" which is distinguished from, and often in tension with, their particular motives and interests. But while private motives and interests are legitimate targets of transformation, two interests which Rousseau thinks all human beings share give his political theory as a whole its character and purpose: the interest in self-preservation and the interest that human beings have in their freedom. Accommodating these fundamental interests, Rousseau thinks, are necessary not only from the perspective of stability, but from the perspective of justice. The interests in self-preservation and freedom, in other words, are not simply interests which affect the plausibility of political arrangements; rather, the way in which these considerations are addressed by a conception of justice affects the very *just-ness* of that conception.¹⁹

Because Rousseau regards freedom and security as the fundamental interests of human beings (about which more will be said in the next section), justice requires that these fundamental interests be embodied in and protected by political theory and political arrangements.

Rousseau writes that man's "first law is to attend to his own preservation, his first cares are those he owes himself" (*SC* 1.2.2). In this regard, Rousseau agrees with Hobbes. But unlike Hobbes, Rousseau maintains that freedom is as important to human beings as their security. Indeed, he asserts that "[t]o renounce one's freedom is to renounce one's

¹⁹ Stuart White reminds us that "a conception of justice simply is, at the end of the day, an account of the shared basic interests of citizens and of the principles that should govern the protection and promotion of those interests" (2003: 25).

status as a man” (SC: 1.4.6). The concern that Rousseau shows for the interests in freedom and self-preservation sits in sharp contrast to the claims of his critics that his theory shows little or no concern for human interests and aims at all and that he desires above all to mold human beings to fit some pre-conceived utopian ideal.

Consider the contrast between Hobbes’s and Rousseau’s political thought. Hobbes holds that individuals’ concern for self-preservation in the state of nature – where there is no “common Power to keep them all in awe” – leads to a “perpetuall and restlesse desire of Power after power, that ceaseth onely in Death” (Hobbes, 1968: 161, 185). To emerge from this terrible condition, in which the life of man is “solitary, poore, nasty, brutish, and short” (1968: 186), Hobbes maintains that each individual would agree to give up her power of self-government – her freedom – to an absolute sovereign on the condition that all others would do likewise. The agreement entails that “every man should say to every man, *I Authorise and give up my Right of Governing my selfe, to this Man, or to this Assembly of men, on this condition, that thou give up thy Right to him, and Authorise all his Actions in like manner*” (1968: 227). In an environment of social interdependence, freedom is held to be irreconcilable with peace and preservation. As such, Hobbes opts for peace at the expense of freedom; for what would freedom consist in, Hobbes might ask, if individuals had time for nothing but the struggle for personal security?

While Hobbes holds that the provision of security and peace requires the alienation of one’s right to self-government, Rousseau maintains that neither security nor freedom can legitimately be cast aside. Recall his belief that “to renounce one’s freedom is to renounce one’s status as a man, the rights of humanity and even its duties” (SC1.4.6). So Rousseau rejects Hobbes’ assumption that freedom and security are necessarily at odds and suggests, rather, that “the essence of the body politic lies in the harmony of obedience and freedom”

(SC3.13.5). The *fundamental problem*, as we have seen, is framed in a way which prevents Rousseau from reaching a Hobbesian solution to the problem of collective security. It says that the form of association which would provide collective security must use means which ensure that “each one, uniting with all, nevertheless obeys only himself and remains as free as before” (SC: 1.6.4). To remain “as free as before”, moreover, each agent need only grant “obedience to the law one has prescribed for oneself” (SC1.8.3).

With those parameters, the fundamental problem is solved by a social contract whereby each member of society “puts his person and all his power in common under the supreme direction of the general will” (SC1.6.9). At first glance, this solution seems to repeat the Hobbesian trade of freedom for security. Indeed, Rousseau writes that the social contract entails “the total alienation of each associate, with all his rights, to the whole community” (SC1.6.6). Moreover, he acknowledges that “what man loses by the social contract is his natural freedom and an unlimited right to everything that tempts him and that he can get” (SC1.8.2). But while the social contract ensures that members receive better protection because it rests on the power of the whole community and because it eliminates the need for power struggles, the very nature of that contract and the security that it provides makes freedom possible. Rousseau does not think that his social contract entails an alienation of freedom for security; rather, he thinks that the social compact is a freedom enhancing or *enabling* agreement.²⁰ Wokler writes that “while for Hobbes, liberty is exchanged for authority in men’s transfer of their natural rights to their ruler, for Rousseau, provided that citizens rule themselves, liberty is *won within* the state rather than protected against it” (1995: 62. My emphasis). To be sure, while Rousseau says that human beings

²⁰ I take the term “freedom-enabling” from Holmes (1988: 195).

lose their “natural freedom”, he maintains that they gain “civil freedom” and “proprietorship” in its place (*SC*: 1.8.2).²¹

In any case, the social compact to which individuals freely agree in order to improve their security entails that members of society are to be ruled by a *general will*. Some critics think that obedience to Rousseau’s general will amounts to an alienation of their private wills and, therefore, their freedom.²² But on Rousseau’s view, to be ruled by the *general will* is to be subject to laws whose enactment rests on an understanding of the common good, which embodies “an equal concern for the good of each citizen” (Cohen, 2001: 6). That “good” is citizens’ shared fundamental interests in security and freedom. In submitting to the direction of the *general will*, individuals submit to a sovereign will which acknowledges their equal worth and which enacts only that legislation which each individual, in a reasonable and reflective way, could agree is consistent with the common good – that is, consistent with their security and freedom. The general will is certainly not, as many critics suspect, a utopian ideal which Rousseau seeks to impose on the members of the social compact; rather, it is an expression of human beings’ own fundamental interests in security and freedom and a mutually agreed upon understanding of how those shared aims can be realized for all equally. Human beings are not clay to be molded or blank slates to be written on; human beings – notably, their capacities and their interests – are the primary inputs for political philosophy and Rousseau’s social contract and general will are an attempt to accommodate those basic features of human life.

²¹ “One must distinguish carefully between natural freedom, which is limited only by the force of the individual, and civil freedom, which is limited by the general will; and between possession, which is only the effect of force or the right of first occupant, and property, which can only be based on a positive title” (*SC*: 1.8.2).

²² Cassirer (1954), Talmon (1952).

II. The Moral Point of View

Rousseau's attempt to transform the social ethos, on the one hand, and accommodate certain fundamental human interests, on the other, might strike us as incoherent. It might seem that he advocates a transformation of the social ethos simply to appease critics in the utopian tradition while making accommodations to certain interests simply to appease realist critics. However, while Rousseau's position is complex, it is not incoherent. The distinction between those interests and motives which require transformation and those which demand accommodation is made from a moral point of view which Rousseau believes is shared by all human beings and which they would recognize if only they would reflect and listen to their "inner voice" or "natural feelings" (*E*: 290, 303). Because that point of view is stable, accessible to all, and universally shared, it provides the basis for reconciling the aims of transformation and accommodation.²³

Indeed, Rousseau's project is coherent in the sense that it embodies a *monological* approach to resolving the tension between justice and the demands of realism. When confronted with possible gaps between features of human beings and the social and political world, on the one hand, and the demands of a candidate conception of justice, on the other, Rousseau advises us to consult the inner voice of conscience to discover the resolution to the apparent gap. Theorizing about and practicing justice, on Rousseau's view, requires agents to engage in a pursuit of truth through isolated, individual reflection. As we shall see, while the monological approach might lead Rousseauian agents to poor conclusions about how to reconcile justice and reality, it nevertheless reveals the possibility of cognitively distinguishing between at least some of the fixed and malleable features of human beings and

²³ On the "inner voice" of conscience, see sections III and IV below.

the social and political world while it simultaneously encourages agents to regard some of their motives and beliefs as possible candidates for legitimate transformation.

In his comments on the *Social Contract* Roger D. Masters offers an important insight into the coherence of Rousseau's project. He writes that

'men as they are' have concerns of 'interest' and 'utility' which 'prescribe' limits to a solution, whereas 'laws as they can be' depend on considerations of 'right' and 'justice' which 'permit' – but only *permit* – of legitimacy. In other words Rousseau will try to adjust 'right' and 'justice' to the *necessary* demands of self-interest" (SC: 133n5).

By emphasizing that Rousseau will adjust principles to the *necessary* demands of self-interest, rather than to *all* demands of self-interest, Masters recognizes that Rousseau does not aim to make accommodations to an objectionable social and political status quo which he thinks is “wicked” and “corrupted”. While Rousseau is committed to reconciling right and interest and thereby acknowledges that political philosophy must respect “the limits of the possible”, he reminds us that these limits “are less narrow than we think. It is our weaknesses, our vices, our prejudices that shrink them” (SC: 3.12.2). So whereas the transition from natural to social man makes principles of justice necessary, Rousseau does not simply construct principles to meet the existing preferences and motives of social man. Rather, he constructs his political philosophy around a few fundamental interests – that is, interests in freedom and self-preservation that human beings have and that it is reasonable for them to have.²⁴

²⁴ Earlier drafts of the dissertation gave the impression that the distinction that matters most to Rousseau is the distinction between interests that are *fixed* and those that are *malleable*. Consequently it appeared as though the aim of a realistically utopian conception of justice should be to accommodate the fixed interests and transform the malleable interests when necessary. Rousseau's view is more nuanced and more appealing than that and I hope to have captured that in the account of Rousseau's view presented here. By fundamental interests, or what I referred to in earlier drafts as “necessary” interests, Rousseau means those interests that must be protected or realized if human beings are to regard themselves as leading valuable and uniquely human lives. Conveniently, Rousseau seems to believe that these *fundamental* interests are also *fixed* interests in the sense of not being capable of elimination from the deep structure of human nature, whereas the objectionable determinate interests about which Rousseau complains he regards as not being the deep structure and are therefore malleable. So, on Rousseau's view, even if certain human beings appear to have, or

And, he constructs his theory with certain features of human beings in mind – namely, that their determinate interests, preferences, and beliefs are malleable; that human beings are capable of critical reflection on their determinate interests, preferences, and beliefs; and that human beings are capable, through such critical reflection, of rejecting or transforming those determinate interests, preferences, and beliefs when necessary. But how do we know which interests are *fundamental*, and therefore require protection in a conception of justice, and which are *not* fundamental, and therefore permit of transformation? Moreover, how does a political theory which attempts to protect certain fundamental human *interests* amount to a realistically utopian theory which offers a solution to the problem of *motivation*?

To answer that first question, consider how, in *Emile*, Rousseau presents an account of how an individual can discover true knowledge despite the corruption of thought by the vices of progress. The account he offers constitutes the core of a *monological* approach to reconciling justice and the demands of realism. Speaking in the voice of a “Savoyard vicar”, he writes that “I do not derive these [principles of conduct] from the principles of the higher philosophy, I find them in the depths of my heart, traced by nature in characters which nothing can efface. I need only consult myself with regard to what I wish to do; what I feel to be right is right, what I feel to be wrong is wrong...Conscience is the voice of the soul, the passions are the voice of the body” (*E*: 298). Indeed, he continues, “[i]t is enough to lead you to distinguish between our acquired ideas and our natural feelings; for feeling precedes knowledge” (*E*: 303).²⁵

believe they have, an interest in acting cruelly, that interest is not one of the fixed interests that are part of the deep structure. To the question, “if cruelty is simply a part of human nature, does it have to be accommodated?”, Rousseau would say that the question does not deserve an answer because the predicate of the question is false – that is, he would hold that cruelty is *not* part of human nature, even if it often appears to be part of the interests or motivational sets of determinate human beings.

²⁵ While the “Creed of a Savoyard Vicar” perhaps best expresses Rousseau’s ideas about the moral point of view, the idea of looking to one’s inner nature for guidance occurred to him as early as the

Notice that the advice given here about looking deeper within oneself parallels Rousseau's "hypothetical history" of the *Second Discourse*. Human beings, as they are, have all manner of determinate and contingent interests, aims, prejudices, and motives. But if we look at human beings in the state of nature – that is, peel back the layers of artificial knowledge and look at the deep structure of human nature – then we begin to see that Rousseau asks us to transform only those interests and motives which are inconsistent with or conflict with the fundamental human interests in self-preservation and freedom. Rousseau's "Creed of a Savoyard Vicar" expresses his conviction that human beings are capable of more than mere instrumental reasoning about how best to satisfy their determinate and contingent interests and aims. He endorses a richer conception of the person which sees human beings as free and equal agents with a capacity for critical reflection on their existing interests and motives rather than mere instrumental actors who regard their preferences and interests as the antecedent and fixed aims of thought and action.

One of the things that makes Rousseau's project interesting for our purposes is that he thinks that critical reflection is best conducted not in conversation or deliberation with other political or moral agents but instead by the isolated individual alone. To be sure, because Rousseau regards social man as corrupted and committed to myriad prejudices, false opinions, and unnecessary preferences, he thinks that guidance about what is moral and just can only be found by drawing back from social man and into the inner depths of natural man. And for knowledge of man as he naturally is, rather than how he is in society, Rousseau holds that inner reflection – monological reflection – is our truest guide. Only by ridding ourselves of social prejudices and preferences can we discover true knowledge about

First Discourse: "O virtue! Sublime science of simple souls, are so many difficulties and preparations needed to know you? Are not your principles engraved in all hearts, and is it not enough in order to

what is moral and just. Thus, the tension between justice and the demands of realism must be resolved, on Rousseau's view, through such monological reflection rather than, say, public political negotiation or deliberation.

Of course, Rousseau's advocacy of a notion of critical reflection on determinate preferences and evaluation of the normative defensibility of various actions and institutions, puts him at odds with realist thinkers who hold that principles of justice are properly understood as solutions to collective action problems where existing preferences and interests are to be regarded as fixed and therefore resistant to transformation. But that doesn't trouble Rousseau because he thinks that a conception of justice, if it is to be a conception of *justice* and not merely a conception of political *order*, must offer an emancipatory critique of the status quo when the status quo is unacceptable from a moral point of view. Existing preferences and interests, he thinks, reflect inequalities and differences in power. Our task is to base political arrangements not on the contingencies of power, but instead on fundamental interests and reason.

If we were to reflect on our natural feelings and listen to our "inner voices", then, Rousseau thinks that we will be able to recognize self-preservation as our first law and freedom as both a fundamental human interest and a basic human capacity. The very act of reflecting on and acting from our natural feelings rather than our social prejudices and artificial interests marks the difference between being a free human being and being a slave to one's passions. With regard to whether Rousseau's projects of transformation and accommodation are coherent, then, the idea that human beings have the power to reflect and to discover through that reflection their most fundamental interests indicates one way in which the projects cohere. That is, those fundamental interests which we find once we listen

learn your laws to commune with oneself and listen to the voice of one's conscience in the silence of

to our inner voices are the proper objects of accommodation whereas all those prejudices or interests which are incompatible with the fundamental interests are legitimate objects of transformation.

So the “inner voice of conscience” gives us two things: In the first place, if we listen closely, the inner voice makes us aware of our fundamental human interests in self-preservation and freedom. And secondly, it gives us a critical standpoint from which we can judge various political arrangements and principles. But the inner voice also gives us a third thing which accompanies the capacity for critical reflection and complements a concern for others’ interest in freedom and self-preservation that we might discover through reason. That third thing, then, is human beings’ natural compassion or pity. Rousseau holds that human beings are inspired by a “natural repugnance to see[ing] any sensitive being perish or suffer, principally our fellowmen” (*D2*: 95). He does not see this as a duty or obligation imposed from the outside; rather, Rousseau regards natural compassion as part of the basic motivational set of human beings. An individual’s “duties towards others,” he writes, “are not dictated to him solely by the belated lessons of wisdom; and as long as he does not resist the inner impulse of commiseration, he will never harm another man or even a sensitive being, except in the legitimate case where, his preservation being concerned, he is obliged to give himself preference” (*D2*: 96).

Thus the inner voice of conscience gives us not only an awareness of our own and others’ fundamental interests and an ability to critically assess proposals for political arrangements or principles of justice in terms of their protection of and contribution to the fulfillment of those fundamental interests, it also makes us aware of our natural compassion – a motivational disposition that encourages in us a concern for the well-being of others. In

the passions” (*D1*: 64).

that sense, the inner voice of conscience does not set out explicit features of a determinate conception of justice nor, for that matter, will it always give us instructions for moral or just behaviour. Rather, the inner voice delivers an awareness of our interests in self-preservation and freedom, reminds us of our natural compassion, and thus provides us with a moral point of view from which we can assess various proposals for political arrangements and principles of justice. And, in light of natural compassion, it ensures that the moral point of view has a “motivational complement” (Cohen, 2001: 54, 57) which allows us to view that moral point of view and the obligations that emerge from it, as an internal feature of our humanity rather than an external constraint on behaviour.²⁶

Whether that monological variant of the realistically utopian approach to the demands of realism – in this case, the problem of motives – is successful or not is a question that will be addressed later in the chapter. For the time being, it is enough to recognize that Rousseau holds that inner reflection is the best strategy for reconciling justice and motivation. If social corruption is the problem – if social corruption is what causes us to regard certain features of an objectionable social and political status quo as either fixed or

²⁶ Sarah Song has suggested that my reading of Rousseau as a defender of the emancipatory potential of critical reflection and reason rests uneasily with Rousseau’s own skepticism of reflection and reason that appears in various passages of his *Second Discourse*. Rousseau writes, for example, that “Reason is what engenders egocentrism, and reflection strengthens it. Reason is what turns man upon himself” (1987: 53). While I can’t offer a detailed reply to Song’s challenge, I think a reply to that concern would move in the following direction: I think that Song is right that there is an important connection, in Rousseau’s mind, between reason and feeling but that feeling should not be confused with the passions. When critical reflection is pursued what we ask is not what the passions demand because the “passions are the voice of the body” (E: 298), rather we ask about what our deepest feelings tell us: “what I feel to be right is right, what I feel to be wrong is wrong...Conscience is the voice of the soul” and “[i]t is enough to lead you to distinguish between our acquired ideas and our natural feelings; for feeling precedes knowledge” (E: 298, 303). When critical reflection is understood as a matter of investigating our true feelings, it can provide the sort of emancipatory potential that I think Rousseau wants it to have in his work. In short, reason or critical reflection may hold the promise of emancipation only if it is oriented to discovering and acting on behalf of human beings’ true feelings rather than their bodily passions and it is reason or critical reflection which helps us to distinguish between bodily passions and the true feelings or conscience that is the “voice of the soul.”

desirable or both – then extracting oneself from the social world through monological reflection must be the best available means to sorting through the distinction between features of human beings and the world that must be accommodated by a conception of justice and those which must be ignored or transformed.

III. Motivational Possibility

Rousseau's accounts of the moral point of view and the psychological capacities of human beings allow him to offer an account of motivational possibility which, as I noted in the introduction, is an important part of a solution to the problem of motivation. But demonstrating motivational possibility requires more than an account of the moral point of view and the psychological capacities of human beings that Rousseau offers. A convincing account of motivational possibility must indicate not simply that the moral point of view is available, but also how that moral point of view motivates or could help motivate citizens to be just. Similarly, an account of the psychological capacities of human beings must be supplemented with an account of how those capacities can be turned towards, or enlisted in the support of, a normatively defensible conception of justice. Indeed, while Rousseau has offered an account of what interests a conception of justice must endorse and protect in order to be normatively defensible, and while he has indicated that human beings have a capacity for critical reflection, we nevertheless still lack an account of how this vision is *motivationally* possible. For instance, if, as Rousseau thinks, a just and legitimate political order depends on our *agreeing* to a social contract to establish that just order and our obedience to only those laws which we give ourselves, then we must be able to regard ourselves as free. That is, if the social contract is an agreement made under duress – whether of force or of the tyranny of the passions – and if laws are imposed rather than

adopted, then the agreement and the laws are not legitimate. But if autonomy is a necessary condition of a just political order, then Rousseau's political theory is realistic only if human beings are free in the appropriate sense.

Again, the moral point of view elaborated in *Emile* – the point of view of the “inner voice” – provides the first step in responding to these concerns. The Savoyard Vicar says that “I was resolved to admit as self-evident all that I could not honestly refuse to believe, and to admit as true all that seemed to follow directly from this; all the rest I determined to leave undecided, neither accepting nor rejecting it, nor yet troubling myself to clear up difficulties which did not lead to any practical ends” (*E*: 278). The idea here is that one can accept as true for *practical* purposes anything which one believes one must accept – for instance, that human beings have freedom of the will if justice requires that – so long as it does not contradict whatever one knows from *theoretical* reason.²⁷ Indeed, when he listens to his “inner voice” for guidance, Rousseau finds, among other things, that human beings have a *moral nature* and that they are *free* to act on principles of that morality. For Rousseau, this means that the motivational demands of the *Social Contract* are not unrealistic because they do not stretch the capacities of human beings beyond the breaking point.

So far, however, that account of motivational possibility is still quite thin. Rather than providing an account of what actively or directly motivates citizens to uphold the political order, here Rousseau has merely provided an account of how the demands of his political order do not demand the motivationally impossible of citizens. Realist thinkers, as we saw in Chapter 1, think that principles of justice will only achieve stability when human beings have the necessary motivation, as well as good reasons, to actively uphold and to live by the principles of justice. Stability of principles of justice is not achieved simply by showing that

²⁷ See Kant (1990b) and Kant (1993).

the demands of justice are not impossible. That human beings are *capable* of doing certain things does not mean that they will actually do those things. For the realist, then, the best way to ensure a fit between principles and motivations is to design principles in the full light of what human beings are already motivated to support – that is, principles which accommodate existing human behaviour, beliefs, and interests whatever those may be. As we have seen above, Rousseau refuses to make accommodations to all facts about human beings because some of those facts are objectionable – indeed, some of them conflict with other more fundamental facts about human beings including the interests in self-preservation and freedom. In the Preface to *Emile* he writes that “People are always telling me to make *practicable* suggestions. You might as well tell me to suggest what people are doing already” (*E*: 2). Thus, he rejects the status quo as having any normative weight.

But from Rousseau’s perspective, the fundamental human interests in freedom and security, along with the capacity to recognize and to act on those interests, gives further motivational strength to his political arrangements. Those fundamental interests are not only accommodated in the political arrangements, they actually *motivate* human behaviour in a certain sense. But is that true? Is it correct to hold that *interests* motivate behaviour or that interests are the same kind of phenomenon as motives? At first glance, we might think that human beings would be or are motivated by their interests. But there is a case to be made against the claim that interests can or should be understood as motives.²⁸ It is easy to imagine instances in which a person has an interest in something, objectively speaking, but that she fails to be motivated by that interest or fails to act in ways which would promote that interest. For instance, a person may have an objective interest in reducing or eliminating her consumption of foods that contain high amounts of trans fatty acids in order to reduce

her risk of heart disease. But that same person might not be motivated to pursue that interest. Instead, she might be motivated by a strong desire to consume rich foods because of the pleasure she receives from the taste of these foods. So recognizing that a person has certain interests – perhaps even certain fundamental interests – does not directly give us an account of that person’s motives. Identification of fundamental interests alone will not provide a solution to the problem of motivation.

Rousseau seems to recognize the distinction between interests and motives and thus the need to develop an even more realistic account of the relationship between justice and motives in *The Government of Poland*. Whereas in the *Social Contract* Rousseau indicates that he wants to reconcile what right demands with what *interest* prescribes, in *Poland* he recognizes that we need also to bring motives into the picture in a more direct way.²⁹ Thus, Rousseau acknowledges in *Poland*, as he does in the *Social Contract*, that “[a] good and sound constitution is one under which the law holds sway over the hearts of citizens” (*P*: 4). But in *Poland* Rousseau also asks “[b]y what means, then, are we to move men’s hearts and bring them to love their fatherland and its laws?” And he answers that it should be “[t]hrough the games they play as children, through institutions that, though a superficial man would deem pointless, develop habits that abide and attachments that nothing can dissolve” (*P*: 4). Elsewhere in that work Rousseau says that we transform human beings into citizens who love justice by tying them “tightly to the fatherland”, by making them patriotic and, at the same time, by tying patriotism to justice and virtue (*P*: 8). The intuition here is that instead of tying the demands of justice to human beings’ fundamental *interests*, which they may fail to perceive, we attempt to tie justice to the beliefs, sentiments, and attachments that already

²⁸ I thank David Estlund for drawing my attention to this problem and for discussion about its significance.

²⁹ On the attempt to reconcile right and interest, see *SC*: 1.0.1.

motivate citizens' behaviour. In that way we can motivate human beings to pursue their fundamental interests and justice without either of those things being felt by them as external constraints.³⁰ If we can successfully tie just behaviour to *internal* motivational possibilities, rather than to external constraint, then the tension between motivation and justice will be significantly, if not entirely, defused.³¹

So Rousseau's account of motivational possibility draws on three ideas: In the first place, he holds that his political arrangements are not too burdensome because they demand of human beings nothing more than what they are capable of doing. That is, the demands do not break the limits of human capacities. Second, positive motivational support for the political arrangements can be found in the fact that they are constructed with a view to accommodating the fundamental human interests in freedom and self-preservation and that motivational complements can be found in the existing beliefs, sentiments, and attachments of potential citizens. And, third, motivational support is found in the idea that the source of those positive motivations is internal rather than external. That is, one need only listen to one's "inner voice", rather than the appeals of others, to gain knowledge of one's most basic interests and to locate the motivational complements which support, but do not alter, the conception of justice. Internal sources of motivation, Rousseau appears to hold, are more likely to provide support to political arrangements than are external sources or sanctions.³²

³⁰ By linking justice to *national* beliefs, sentiments, and attachments, Rousseau raises serious questions about global justice and motivational compliance. Are there any existing beliefs, sentiments, and attachments at an international level which might be enlisted to solve the tension between motivation and global justice? If not, does this suggest that global justice is not possible? Or that it is something altogether different than justice at the level of the nation-state? I intend to pursue some of these questions in future work.

³¹ It is important to note here that in *Poland* Rousseau does not seek to alter the content of justice or virtue while trying to realize justice and virtue in the Polish constitution and state; rather, he merely suggests that justice and virtue can become effective in this real political society if we can tie its demands to the already existing, and motivationally efficacious, beliefs, sentiments, activities, and attachments of the Poles.

³² This idea will be further developed throughout the dissertation, especially in the chapters in Part II.

IV. “Real Interests”, Coercion, and Motivation

While Rousseau offers some instructive ideas about how motivational concerns should be dealt with in a theory of justice, the three features of his solution to the problem of motivational possibility present some new difficulties. Two difficulties in particular should be considered: The first, which I take up in this section, is that the move to identify the “fundamental interests” of human beings prior to democratic deliberation risks making citizens the objects rather than subjects of political discourse and, consequently, unable to make contact with the sources of motivation on which Rousseau’s theory depends. The second, which is taken up in section V, is that Rousseau’s move to find part of the account of motivational possibility in the “inner voice” of human beings, and his claim that all human beings have *fundamental* interests in freedom and self-preservation, collide with the range of reasonable pluralism which, as I shall argue, is a condition which must be accommodated by a conception of justice. Thus, while Rousseau makes some progress in resolving the tension between justice and motivation, that progress comes at the expense of a satisfactory solution to the broader concern of the demands of realism more generally. That is, because Rousseau’s solution to one demand of realism – i.e., motivation – fails to adequately address another demands of realism – i.e., the range of reasonable pluralism – the project as a whole offers an incomplete approach to the tension between justice and the demands of realism more generally.

Williams notes that Rousseau makes a distinction between the “real interests” which *he* thinks human beings have and “the interests *they* think they have.” That distinction, Williams continues, “has generated a vast literature, and almost an equal amount of suspicion...[because] an appeal to people’s real interests is often deployed as a reason for

coercing them contrary to their ‘apparent’ (that is to say, perceived) interests” (1985: 40). Indeed, some critics of Rousseau’s theory hold that the “real interests” approach prohibits citizens from having other private interests that differ from their alleged “real interests” and, moreover, they hold that this approach legitimizes the use of coercion to achieve some utopian aim. The critics worry that the transformative aims of Rousseau’s political theory would treat human beings as mere means to some utopian end. To put it another way, these critics hold that Rousseau’s political theory does not go far enough in accommodating the interests and motives which citizens might legitimately develop when freedom is properly respected.

Cassirer pursues this line of criticism when he argues that the transformative aim of Rousseau’s political thought prevents individuals from having private wills and interests of their own. And that means that individuals who are party to Rousseau’s social contract will find themselves in an oppressively conformist state. He argues that,

The *Contract social* proclaims and glorifies a completely unbounded absolutism of the state. Every particular and individual will is shattered by the power of the *volonté générale*. The very act of joining the state signifies the complete renunciation of all particular desires. Man does not give himself to the state and to society without giving himself completely to both. We may speak of a real “unity” of the state only if the individuals are merged in this unity and disappear in it (1954: 52).

Similarly, J.L. Talmon, in *The Rise of Totalitarian Democracy*, claims that Rousseau’s “aim is to train men to ‘bear with docility the yoke of public happiness’, in fact to create a new type of man, a purely political creature, without any particular private or social loyalties, any partial interests” (1952: 42). Talmon repeats the often heard assertion that “Rousseau’s ‘general will’...became the driving force of totalitarian democracy” (1952: 6). Indeed, in seeking to transform preferences and motives which would be regarded as objectionable from the point of view of an allegedly utopian society, Rousseau seems not to take seriously his own

statement that he will take people “as they are.” As a result, his political theory appears hostile to one of the values that he claims to endorse – namely, freedom. So one problem with the notion of “real interests”, then, is that it seems to legitimize coercive measures that would bring citizens in line with some fundamental interests that they are said to have, even if they fail to perceive those interests themselves. And while those interests might seem worthwhile in their own right, coercion in service of perfectionist aims is objectionable because it fails to treat human beings as free.

These versions of the “real interests” objection, however, rest on a misreading of the *Social Contract*. Rousseau does not say that individuals cannot have private wills; he says only that private wills, when they conflict with the general will, must yield. The social contract and the *general will* do not seek to eliminate private or non-shared interests, but rather to coordinate the shared interests in security and freedom. Nevertheless, it is true to say that even if private wills are permitted, they are required to yield to the general will when the two come into conflict. And it’s that yielding requirement which concerns critics like Cassirer and Talmon, especially when considered alongside Rousseau’s claims that human beings must be “forced to be free” and that “the less the individual wills relate to the general will, that is to say customary conduct to the laws, the more repressive force has to be increased” (JC: 3.1.13). Even if Rousseau thinks that necessary coercion simply brings human beings closer to satisfying their fundamental interests in freedom and security, the very use of coercion would violate Rousseau’s own desire to offer a politically *legitimate* alternative to the Hobbesian trade of freedom for security.

Yet, as the section of the *Social Contract* on the Legislator shows, Rousseau holds that the transformative project cannot be achieved through force because that would conflict with men’s freedom to obey only those laws which they give themselves. The legislator,

Rousseau writes, “must have recourse to another order of authority, which can win over without violence” (SC: 2.7.9). So the problem with the idea of “real interests” is not that it legitimizes coercive means to achieve the necessary transformations because coercion is ruled out by Rousseau as incompatible with human beings’ “real interest” in freedom. Cassirer and Talmon recognize a tension in Rousseau’s theory between the interests that people perceive themselves to have and the fundamental interests which Rousseau alleges that they all have, but they are wrong that the tension necessarily entails coercive measures, on the one hand, and a prohibition on private wills, on the other.

Still, there is an unresolved problem with the notion of “real interests” and its role in Rousseau’s transformative project. Even if Rousseau can avoid the charge that he endorses coercion, there is still the question of how his legislator can get people to agree to those laws which support their “fundamental interests” in security and freedom. In Rousseau’s view, the gap between citizens’ real interests and their perceived interests or between the general will and their private wills is primarily the consequence of citizens’ own rational or informational shortcomings. All citizens share the fundamental interests in freedom and security even if they fail to recognize that they have those interests. Men as they are, Rousseau seems to think, have been corrupted by unjust social and political institutions and so cannot always see what laws would serve their fundamental interests. Indeed, although Rousseau believes that human beings have a capacity for reflection and can perceive their real interests under the right conditions, he is pessimistic about the possibility that such critical reflection can occur on its own without assistance of some kind. While those socially corrupted human beings might be unable to perceive their real interests *as individuals*, perhaps through democratic deliberation, the pooling of information, and the collective assessment of that information they would be able to perceive those fundamental shared interests *as a*

democratic community. That is, if Rousseau offered a political theory in which citizens could deliberate with each other as free and equal citizens then those citizens who failed to recognize their real interests might be brought to recognize them in a non-coercive way.³³

However, Rousseau seems to eliminate the deliberation option: “If there were a people of Gods,” he writes, “it would govern itself democratically” (*SC*: 3.4.8). “[I]f there were no different interests, the common interest, which would never encounter any obstacle, would scarcely be felt. Everything would run smoothly by itself and politics would cease to be an art” (*SC*: 2.3.note). But, Rousseau continues, “such a perfect government is not suited to men” (*SC*: 3.4.8). So Rousseau’s theory, even as it eschews coercion and therefore escapes the criticisms of Cassirer and Talmon, also appears to eschew a deliberative version of democracy which leaves the problems of rational and informational deficiencies about “real interests” unsolved. And, in many ways, this problem points to one of the main difficulties encountered when one adopts a *monological* approach to resolving the justice and realism tension. That is, by relying on inner reflection and conscience to discover an alleged moral point of view, agents are deprived of the better information and critical encounters that they could find in a more deliberative and public attempt to work through the justice and realism tension.

As Williams recognizes, however, the most troubling feature of the real interests problem is not that citizens exhibit rational and informational deficits when it comes to their alleged “real interests”. Rather the problem is that, even if those real interests were *known* by citizens, the gap between those real interests and their perceived interests means that citizens

³³ “A well-constituted deliberative forum,” write Gutmann & Thompson (2004), “provides an opportunity for advancing both individual and collective understanding. Through the give and take of argument, participants can learn from each other, come to recognize their individual and collective misapprehensions, and develop new views and policies that can more successfully withstand critical scrutiny” (2004: 12).

still might not *feel* those interests as their own. “The most significant question about real interests arise,” Williams writes, “when what is wrong with the agent goes beyond lack of information or mere rationality...and affects the desires and motivations from which he deliberates; or, again, when what’s wrong with the agent is that he will not believe something that he rationally should believe” (1985: 41). Again, that problem might be overcome through democratic deliberation: If knowledge of, and laws that reflect, the real interests were worked out in a deliberative fashion and citizens made judgments democratically, then they might be motivated by the fact that they gave themselves the laws to which they are subject. But just as the Legislator cannot use coercion to get people to agree to laws which are in their best interests, in Rousseau’s theory, neither can the Legislator use reason because people do not yet have the virtuous “social spirit” which would help them to recognize and accept good laws. “The social spirit, which should be the result of the institution,” Rousseau writes, “would have to preside over the founding of the institution itself; and men would have to be prior to laws what they ought to become by means of laws” (*SC*: 2.7.9).

By appealing to a notion of real interests which a conception of justice must accommodate and towards which citizens must be transformed, and by rejecting coercive and democratic means to resolve some of the difficulties with that approach, Rousseau appears to have failed to solve the problem of motivational possibility. That is, by simply pointing out that human beings have certain fundamental interests in freedom and security, Rousseau appears to have failed to show how human beings could be motivated to support those interests when they conflict with their other private, and more immediately perceived, interests. Even when Rousseau tries to link the demands of justice and the pursuit of the fundamental interests to national beliefs, sentiments, and attachments, as he does in *Poland*, it is not clear why we should expect those motives to outweigh other, perhaps objectionable,

motives that agents have or, as we will consider presently, the other beliefs to which agents might be committed. Thus, Rousseau's skepticism of more deliberative, participatory versions of democracy is especially troubling given the way that those arrangements might help to overcome the rational, informational, and motivational deficits that vex a monological variant of realistic utopianism. Moreover, the rejection of deliberative democracy which has the potential to resolve the problem of motivational possibility is particularly puzzling given Rousseau's own claim that human beings are free and that legitimate political arrangements must be those to which citizens could agree.

V. Reasonable Pluralism and the "Inner Voice"

The other main problem with Rousseau's monological approach to resolving the justice and realism tension is that he fails to acknowledge the demands imposed by the range of reasonable pluralism and the need to reconcile justice to that range.³⁴ The moral point of view which helps to distinguish between legitimate objects of transformation and the objects of accommodation is not shared by all citizens in a diverse society and that creates serious difficulties for Rousseau. One is a practical problem that arises from the tension between the moral point of view and the fact of pluralism. The other is a more serious problem – a problem of justice – which arises from a tension between the moral point of view and the range of *reasonable* pluralism. I will address each of these problems in turn.

1. Pluralism and Stability

In the first place, Rousseau's turning to the "inner voice" of the subject as the appropriate moral point of view suffers from a practical problem. Recall that Rousseau believes that the

moral point of view is internal rather than external. That is, one need only listen to one's "inner voice", rather than the appeals and concerns of others, to gain knowledge of one's most fundamental interests and one's sense of natural compassion. That meant that the transformations of preferences and interests to which citizens might be subjected could find their justification in interests that citizens already have and would recognize they have if only they peeled away their accumulated prejudices and biases. Moreover, because the moral point of view is internal rather than external, Rousseau thinks that he is offering a more compelling account of motivational possibility. That is, the motivation to endorse and abide by the demands of justice is something an agent already has rather than something imposed externally. In short, Rousseau thinks that justice is best determined, and the problem of motivation is best solved, by adopting a monological version of the realistically utopian project.

But that moral point of view, which says that when we listen to our inner voice we discover that we are naturally and fundamentally free beings, that we are interested in our self-preservation, and that we are naturally compassionate, is not shared by all citizens or potential citizens as a matter of fact. Some citizens might think that the *source* of morality lies elsewhere and that the *content* of that morality excludes the idea of natural freedom. Indeed, there are religious views which hold that human nature and conscience are not the primary sources of morality; rather, the moral point of view is the point of view found in sacred texts or that of a transcendent and omnipotent deity. Contemporary political societies are characterized by a pluralism of moral and religious worldviews which means that many citizens will not regard themselves as having reason to endorse Rousseau's moral point of view or the conception of political right that rests on that point of view. Moral points of

³⁴ Cohen (1993), Rawls (1996). How we should go about accommodating that range in a conception

view and conceptions of political right which appear to conflict with their beliefs might simply be rejected by those citizens in order to preserve and protect the integrity of their worldviews.

When the fact of pluralism obtains, then, it creates great difficulties for any view which bases principles of justice or political arrangements on the language of legitimacy and on a point of view which is not shared. Because Rousseau thinks his political order must be established by an act of free will for the purposes of protecting individuals and their freedom, it fails to take seriously the objections of those individuals who do not accept freedom as a moral ideal or who disagree that the moral point of view is located in human conscience. That failure to account for the objections of *all* citizens creates unfair challenges for those citizens who are otherwise motivated to make sacrifices to uphold the political arrangements. Even if the political arrangements achieved some measure of stability in the face of pluralism, motivated citizens would face the possibility of exploitation by the unmotivated.³⁵ That is, while those citizens who are motivated by or who regard themselves as having reason to endorse Rousseau's moral point of view make the necessary changes to their other preferences and beliefs, citizens who regard themselves as *not* having reason to endorse Rousseau's moral point of view might continue to make demands on the basis of their pre-political, unevaluated preferences and beliefs. Without a shared point of view to resolve these disputes, conflicts would have to be resolved through bargaining and the

of justice will be the subject of chapter 5.

³⁵ In written comments on earlier drafts and in private conversations, Joe Heath has referred to this as the "exploitation of the moral by the immoral". Those citizens who are moral – those who sacrifice their interests and preferences in the interests of fairness and equality for all – are exploited by the immoral – those citizens who neither make such sacrifices and who, moreover, can gain some additional benefit by preying on the moral behaviour of others. I have chosen to refer to this instead as the exploitation of the motivated by the unmotivated in part because exactly what is "moral" is part of the dispute.

outcomes of those bargains would likely reflect differences in power rather than fair agreements.

Thus, we begin to see that a monological variant of the realistically utopian project has difficulty once we recognize that a conception of justice is not for individuals alone, but rather constitutes the normative basis for how we are to live and act *together*. And it has difficulty once we recognize that conceptions of justice must confront a social reality which is characterized by moral and religious diversity rather than homogeneity. In light of the fact of our *social* and *plural* existence, then, a monological approach to justice and the demands of realism seems less promising insofar as it blinds us to significant and theoretically relevant features of social and political life.

One might think that Rousseau can avoid this difficulty by appealing to his transformative aims. Recall that Rousseau is prepared to set up institutions which will create more virtuous citizens – that is, citizens more willing and able to reflect on and to reject when necessary their non-fundamental interests, aims, and preferences when they conflict with fundamental interests. Indeed, Rousseau might simply hold that moral and religious views which do not identify or fit with his moral point of view, or agents who fail to locate the moral point of view in their own “inner voice” or conscience, are simply wrong and therefore subject to criticism and transformation. Diversity or the fact of pluralism might be regarded as the consequence of mistakes in reasoning about moral or religious ends or the result of coercion by moral and religious authorities rather than un-coerced reflection and conviction. We might even agree with Rousseau that some moral and religious views which do exist are objectionable and therefore should not affect the content of a conception of justice and political arrangements. Some religious views, for example, endorse race and

gender hierarchies and are clearly incompatible with what reasonable people would regard as fair.

While Rousseau might be right not to capitulate to such views, the stability of his political arrangements will be in jeopardy if too many of the moral and religious views with which he disagrees are excluded.³⁶ Indeed, as numerous views are excluded and marked for transformation, it becomes hard to see how agreement and legitimacy continue to play the roles in his political theory that he wants them to play. At best, Rousseau's normative vision would be paternalistic; at worst, it would become the objectionably utopian and totalitarian ideal about which his critics complain. Nevertheless, while this concern about stability in the face of pluralism is a difficult challenge – and one which I don't fully address until later in the dissertation – the challenge is, ultimately, a practical challenge. We want agreement, to be sure, but not an agreement which makes concessions to *objectionable* moral and religious worldviews.

2. *Reasonable Pluralism and Justice*

Once we move from the fact of mere pluralism to the fact or range of reasonable pluralism, any thought that Rousseau can somehow rescue his moral point of view should arouse skepticism. Whereas the fact of pluralism simply describes the fact that there are a variety of different, incompatible, and perhaps conflicting moral and religious views in a society, the range of *reasonable* pluralism describes a more challenging condition. The idea of a range of reasonable pluralism entails, first, that a subset of the different moral and religious views are compatible with fair political arrangements even while they reject the moral point of view

³⁶ I am not saying here that Rousseau should accommodate such views, only that the stability of his political arrangements will suffer if they don't connect in some way with the motives and sources of motivation of diverse moral and religious views.

which motivates Rousseau's human beings. Moreover, the range of reasonable pluralism entails that there is a set of reasonable views held by people who are prepared to agree to fair terms of social cooperation even with others who hold different worldviews. Finally, the diversity of comprehensive and incompatible, albeit reasonable, religious, moral, and philosophical doctrines is the "natural outcome of the activities of human reason under enduring free institutions" (Rawls, 1996: xxvi).³⁷

As Rawls puts it,

The diversity of reasonable comprehensive religious, philosophical, and moral doctrines found in modern democratic societies is not a mere historical condition that may soon pass away; it is a permanent feature of the public culture of democracy. Under the political and social conditions secured by basic rights and liberties of free institutions, a diversity of conflicting and irreconcilable – and what's more, reasonable – comprehensive doctrines will come about and persist if such diversity does not obtain already (1996: 36).

The doctrines are *reasonable* – or, more precisely, the people who hold such doctrines are reasonable – when they acknowledge that the very rights and liberties protected by just institutions which facilitated the development of their own doctrine permits the development of others and when they regard that as an acceptable consequence of political freedom. Resolving conflicts and coming to agreements, then, have to be cast not in terms of what an independent moral ideal advises us to do; rather, reasonable people seek out agreement that rests on reasons and conclusions that other reasonable agents could endorse. And that means that reason-giving has to be political – i.e., it has to refer to the terms of social cooperation, treating people as *citizens* as politically free and equal and reasonable – rather than moral – i.e., referring to some comprehensive view about the person or nature, treating people as clay to be shaped according to the demands of the moral ideal.

³⁷ See also Cohen (1993).

The problem for Rousseau, then, is not simply that his political theory will be *unstable* under conditions of reasonable pluralism, but that it will fail to accord all parties to a social contract the respect that is owed to politically reasonable citizens as a matter of *justice*. To be sure, insofar as reasonable pluralism is the natural outcome of free practical reason under just institutions and insofar as the citizens who hold such doctrines are prepared to agree to fair terms of social cooperation with others who don't share their moral or religious worldview, accommodating that fact is necessary as a matter of justice and not simply as a matter of practicality.³⁸ But Rousseau's political theory requires the transformation of persons into citizens who share his moral point of view and thus requires otherwise reasonable citizens to cast off their moral and religious views to become free and equal citizens in Rousseau's political community.

What is striking about Rousseau's view is that while on the one hand he acknowledges that founding a state on a particular religion is "more harmful than useful", he nevertheless maintains that his own doctrine of natural freedom admits of the same difficulties that he seeks to avoid by eschewing particular religions. In the section on Civil Religion in the *Social Contract*, Rousseau writes that "[e]verything that destroys social unity is worthless. All institutions that put man in contradiction with himself are useless" (SC: 4.8.17). To be sure, while he endorses a "civil profession of faith" whose dogmas are to act as "sentiments of sociability without which it is impossible to be a good citizen or a faithful subject" (SC: 4.8.32), Rousseau allows that a variety of existing religious opinions will be compatible with that civil faith. It seems as if Rousseau recognizes that some disagreement is reasonable and that it should be accommodated rather than transformed. But that view appears only to apply to a small subset of Christian religious doctrines and not to Catholicism, other non-

³⁸ This argument is developed much more fully in chapter 5.

Christian doctrines, or even non-religious comprehensive views. A subset of Christian doctrines, it would seem, is compatible with Rousseau's vision because, he suggests, those doctrines are compatible with his idea that human beings are at base free – that God made them so. In *Emile* he responds to those who complain that God does not prevent human beings from doing evil by arguing that such a view fails to acknowledge that human beings are free to do good or evil because God made people free: "Providence has made him free that he may choose the good and refuse the evil" (*E*: 292). So when Rousseau says that those "institutions that put man in contradiction with himself are useless" he means those institutions which conflict with the idea that human beings are essentially free and that God made them so.

But once one recognizes human beings as free and once one sets up institutions which acknowledge and try to promote that freedom, people will develop other comprehensive doctrines which conflict with the idea that they are free in the Rousseauian sense while, at the same time, those people are able to acknowledge other citizens as *politically* free and equal. Rousseau himself acknowledges that human beings are beset by what Rawls calls the "burdens of judgment" which lead them to endorse different comprehensive doctrines under institutions which protect freedom and liberty.³⁹ The fact and the idea of a range of *reasonable* pluralism create a dilemma for Rousseau's political theory. Rousseau is unable to give a persuasive account of how individuals who would endorse different comprehensive doctrines could come to agree to the original social contract in the first place. Those people who disagree that human beings are naturally and essentially free and who disagree that the moral point of view is located in human beings' inner conscience would refuse to agree to a

³⁹ In *Emile* Rousseau writes: "I need not think myself infallible; my opinions, which seem to me true, may be so many lies; for what man is there who does not cling to his own beliefs; and how many

social contract and political arrangements which accepted those things as true. Moreover, they would refuse on reasonable grounds – grounds that Rousseau must acknowledge are reasonable given that they could emerge under the institutions he favours (i.e., institutions which protect freedom); conditions he acknowledges (i.e., the burdens of judgment); and given that the people who endorse those other comprehensive doctrines are capable of recognizing other citizens as free, equal, and reasonable. So long as the *fact of reasonable* pluralism obtains, there is a relevant and unobjectionable demand of realism which a realistically utopian political theory must acknowledge and accommodate – especially one that recasts the tasks of political theory in terms of legitimacy rather than perfectionism. And so long as we can conceive of a *range* of reasonable pluralism – even if all possibly reasonable views are not instantiated in social and political life at present – then a conception of justice will have to wrestle with just how to reconcile itself to that condition.

VI. Conclusion: The Perspective of a Realistic Utopia

Even as Rousseau fails to accommodate the range of reasonable pluralism in his political theory – and thereby fails to accommodate a feature of human beings which demands accommodation as a matter of justice and not mere prudence – he instructs us on the questions and concerns that a realistically utopian theory should address. Moreover, he offers a few ideas which might guide the development of a more attractive approach to the tension between justice and realism. Indeed, his notion of a moral point of view, albeit problematic for reasons described above, gives us a way to think about which demands of realism should be accommodated and which should not be accommodated in a conception of justice. His conception of human beings as free, equal, and capable of critical reflection

men are agreed in everything? The illusion which deceives me may indeed have its source in myself,

on their determinate preferences, interests, and motives – albeit incompatible with certain moral and religious views – points towards an idea which helps to solve the problem of motivational possibility. And, finally, his idea that the political institutions which help to shape citizens' sense of justice, at the same time, depend on citizens' free consent for their legitimacy and motivational stability, is another idea that will play a central role in a more compelling reconciliation of justice and realism. So just how should these ideas be improved upon? To begin to answer that question we turn, in the next chapter, to John Rawls's attempt to offer another variant of the realistically utopian approach to the justice and realism tension.

but is God alone who can remove it" (*E*: 308).

3

Rawls: Hypothetical Constructivism

[P]olitical philosophy is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition.

- John Rawls (*LP*: 11)⁴⁰

[T]here is a question of how the limits of the practicable are discerned and what the conditions of our social world in fact are; the problem here is that the limits of the possible are not given by the actual, for we can to a greater or lesser extent change political and social institutions, and much else.

- John Rawls (*JF*: 5)

Like Rousseau, John Rawls regards the tension between justice and realism as significant and in need of a solution. Whether a conception of justice can achieve motivational stability, for example, is treated not as a mere practical issue but as a concern which matters to the very normative attractiveness of the conception of justice itself. But again, like Rousseau, Rawls does not want to offer a conception of justice which amounts to a mere rational bargain among political actors fully aware of their determinate preferences, motives, and beliefs. Instead, he wants to articulate principles of justice which are insulated from agents' objectionable preferences, motives, and beliefs – that is, principles which maintain a critical distance from the social and political status quo and its resource and power inequalities. As Rawls says, “[j]ustice is the first virtue of social institutions, as truth is

⁴⁰ Rawls's works are cited in this chapter as follows: *CP*: *Collected Papers*, Samuel Freeman ed., (Cambridge: Harvard University Press, 1999); *JF*: *Justice as Fairness: A Restatement*, Erin Kelly, ed. (Cambridge: Belknap, 2001); *LP*: *The Law of Peoples* (Cambridge: Harvard University Press, 1999); *PL*: *Political Liberalism* (New York: Columbia University Press, 1996); *TJ*: *A Theory of Justice. Revised ed.* (Cambridge: Belknap, 1999); and *IPRR*: “The Idea of Public Reason Revisited” in Rawls, *The Law of Peoples*.

of systems of thought...[L]aws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust” (*TJ*: 3).

In that sense, Rawls seems to align himself with the utopian tradition of political philosophy which seeks to construct or articulate principles while bracketing pluralism and various motivational difficulties that arise when principles are applied to real societies. Nevertheless, Rawls also thinks that political philosophy should have a practical orientation. “Conceptions of justice,” he writes, “must be justified by the conditions of our life as we know it or not at all” (*TJ*: 398). He wants to find a solution to the problems of motivational possibility and doctrinal pluralism by gaining public agreement on the principles of justice among citizens with diverse interests, conceptions of the good, and reasonable comprehensive doctrines about human ends and ways of life. With these dual aims of critique and practicality at the centre of his work, Rawls defends a realistically utopian conception of justice which in part ties the justification of principles to their legitimacy and, in so doing, indicates that concerns about motives and pluralism are relevant to the content of justice. Rawls’s *hypothetical constructivist* variant of realistic utopianism holds that principles of justice should be the outcome of a process of reflection on relevant reasons. In some cases the demands of realism enter into that reflection as relevant reasons – as concerns that other reasonable individuals could regard as demands that require some kind of accommodation in principles of justice. At the same time, however, because the approach is *hypothetical* – because it requires that agents bracket their determinate preferences, motives, and beliefs – Rawls’s version of realistic utopianism faces a democratic challenge to which, as we shall see, he has difficulty responding.

In this chapter, then, I focus in particular on the way in which motivational concerns enter into Rawls's attempt to develop an attractive conception of justice.⁴¹ I investigate whether Rawls's attention to the so-called "strains of commitment" (*TJ*: 153-4) leads him to make accommodations to objectionable motives or motivational concerns. I argue that, in light of Rawls's later revisions of the role of the primary goods and the conception of the person, attention to the strains of commitment does not force concessions to objectionable motives. However, a defense of Rawls's use of the strains of commitment meets resistance from G.A. Cohen (2000, 2003) who argues that justice as fairness in fact does make objectionable concessions to existing motives in the content of the difference principle. These concessions, Cohen alleges, are the upshot of Rawls's requiring adherence to justice in the basic structure of society rather than in the choices and ethos of citizens themselves. While promising replies to Cohen's argument have highlighted the fact that, contrary to Cohen's claims, Rawls does think that transformation of an objectionable social ethos is necessary, none of these seem to have captured the centrality of an ideal of democratic legitimacy to Rawls's conception of justice as a whole. To meet Cohen's criticism, then, I will maintain that locating justice in the basic structure is one way in which a conception of justice accommodates citizens' interests in liberty. Thus, while adjustments are made to the conception of justice to accommodate a particularly important motive – namely, liberty – those adjustments should be regarded not as concessions to some objectionable feature of the world, but instead as a matter of a conception of justice recognizing and endorsing a motive – i.e., a concern for liberty – which any normatively defensible conception of justice must do. How far we should go in accommodating specific manifestations of, or claims

⁴¹ We will deal with the other demand of realism – pluralism – in Chapter 5.

about, liberty and freedom, however, is a question that should be resolved by citizens themselves once the content of principles of justice are settled.

I begin with an examination of Rawls's original position and veil of ignorance which capture the core of his *hypothetical constructivism* – i.e., the idea that principles should be insulated from objectionable motives and that agreement should be reached from an initial situation of equality. Section II asks whether Rawls's use of a “strains of commitment” requirement allows certain objectionable motives to slip into the design of principles of justice and, in section III, I consider G.A. Cohen's objection to Rawls's theory in Cohen's *If You're an Egalitarian, How Come You're So Rich*.⁴² To answer Cohen's charge, I argue that once we have a proper understanding of the revised role that Rawls gives to primary goods and the conception of the person in justice as fairness, we begin to see how the strains of commitment approach addresses the problem of motivation in an appropriate way (Section IV). The chapter concludes, in section V, with some preliminary thoughts about how the content of a conception of justice for a democratic society can and should accommodate the democratically defensible motives of democratic citizens. Those thoughts will get us closer to the idea, more fully developed in the final chapter, that a conception of justice must address the demands of realism because of the relationship between realism, legitimacy, and justice.

⁴² It's useful to distinguish between two lines of criticism that Cohen aims against Rawls's justice as fairness. In his earlier work on Rawls, Cohen's concern is that the argument for the difference principle relies on a concession to incentive-seeking behaviour on the part of talented citizens (Cohen, 2000). It is that criticism that I take up in this section. The other line of critique is introduced and developed in Cohen (2003) and Cohen (2003b). In those works, Cohen takes issue with Rawls's statement that “Conceptions of justice must be justified by the conditions of our life as we know it or not at all” (*TJ*: 398). Cohen argues instead that facts cannot “ground” principles and that Rawls is misguided about the principles motivating his own work when he considers what facts

I. Agreement, Motivation, and Justice as Fairness

The main idea behind justice as fairness, Rawls writes, “is that the principles of justice for the basic structure of society are the object of [an] original agreement” (*TJ*: 10). Of course making the content of justice depend on some kind of agreement raises questions about how that agreement is to be reached and what the contractors are allowed to know about themselves. Does Rawls’s reliance on the idea of a social contract prompt him to adjust the content of principles of justice themselves in order to elicit agreement which is necessary for the legitimacy and stability of the conception of justice? Indeed, does Rawls’s commitment to a Rousseauian ideal of self-government – of obedience only to those laws one gives oneself – allow citizens’ existing motives to shape the content of principles in objectionable ways?

Social agreements and the idea of moral arbitrariness

Rawls says that acceptable principles will be those “that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association” (*TJ*: 10). While Rawls wants to follow in the social contract tradition and thereby ensure that his conception of justice embraces an ideal of free and self-governing citizens, he departs from much of that tradition in seeking a fairer and more equal initial position for all contracting parties. He holds that certain features of human beings are “arbitrary from a moral point of view” (*TJ*: 14) and therefore should not affect the content of principles of justice. Features such as race, class, sex, and the distribution of natural abilities are the outcome of a natural lottery. And because the outcomes of that lottery are dictated by chance and luck, rather than choice, we should not

about human beings and society require accommodation in order for a conception of justice to be

be permitted to prosper from, nor doomed to suffer because of, those particular features that we end up with.

The “moral point of view” from which we determine whether a feature is relevant or not, Rawls thinks, is implicit in our everyday common sense moral awareness, but is rarely expressed with the sort of philosophical precision which would be necessary in order to identify defensible principles of justice. Thus, Rawlsian political philosophy starts from certain “settled convictions” of our shared political morality – such as that slavery is wrong – and tries to construct general principles of justice which rely on moral principles implicit in our common sense morality. Working back and forth between settled convictions and general principles, justice as fairness aims to achieve reflective equilibrium with our common sense morality. With a reconstructed moral point of view as a guide, then, Rawlsian justice distinguishes between those motives, among other things, which are relevant and those which are irrelevant for the content of principles of justice.⁴³ This constructivist approach, then, relies on the reflective capacities of the agents to whom the principles will eventually apply. That is, the legitimacy of the hypothetical constructivist variant of realistic utopianism relies on the ability of agents to abstract from their own preferences, motives, and beliefs and to think through the requirements of the intuitions and ideals embedded in the public political culture of which they are members. But, because the procedure is *hypothetical*, it means that those reflecting on principles of justice and the reasons which might be relevant in their construction, must make a leap of theoretical abstraction – that is, they must imagine

just.

⁴³ But while Rawls says that the moral point of view is located in our implicit common sense morality, it seems more accurate to say that the moral point of view is implicit in our public reasoning about political and moral questions. That is, what we regard as relevant and irrelevant is implicit in our settled convictions *because* those settled convictions represent claims in public reasoning to which all have or could agree. This idea points towards the Habermasian foundation of a moral point of view which will be considered in chapter 4.

themselves as agents who are rational and reasonable, but who have no particular preferences, motives, and beliefs. An agreement about principles of justice, if one can be reached, is thus not an actual agreement between fully self-aware agents or citizens; rather, principles of justice are legitimate, if they are, because reflective agents can imagine that such principles could be the product of rational and reasonable agents situated behind a veil of ignorance.

From the moral point of view, then, Rawls holds that a host of features of existing human beings and the social and political world have no relevance for the initial agreement on principles of justice. Consider, for example, the relevance of a pre-contractual distribution of property for a conception of justice. In Locke's theory the distribution of property prior to the social contract is not an appropriate object of criticism if holdings were acquired legitimately in the state of nature.⁴⁴ That is, so long as it was acquired in the right way – either by initial acquisition subject to the conditions of non-spoilage and leaving enough and as good for others, or by consensual transfer from one party to another – then the distribution of those holdings among citizens is legitimate. But legitimate acquisition is compatible with unequal holdings which means that some of the parties to Locke's social contract will be motivated to argue for and adopt principles and institutions which would protect existing inequality rather than, say, ensuring a more equal distribution of resources. If an agreement is reached among contractors who have unequal holdings and who are aware of their unequal holdings then Locke might succeed in reconciling what justice demands with the motives of some, but not all, citizens. In short, the content of Locke's agreement or conception of justice would be shaped to suit the interests of a landed class rather than all citizens. But if *agreement* is what Locke is after, then it is hard to see why the

⁴⁴ For the account of legitimate acquisition see Locke (1980: chapter 5).

landless classes would agree to arrangements in which discussions about the legitimacy of the existing property distribution would not be permitted. Indeed, one of the main purposes of a social contract on Locke's view is to ensure the protection of property that individual contractors had acquired in the pre-contractual state of nature.⁴⁵

But taking those features of the pre-contract status quo as fixed features which contractors must accept is objectionable from Rawls's point of view because that status quo was the product not of transactions under just institutions, but instead influenced by the contingencies of unequal power, order of discovery, and order of birth, and maintained by unregulated force rather than legitimate and just law. The moral point of view implicit in our common sense morality and settled convictions, Rawls thinks, would regard distributions that reflect power inequalities and order of discovery or birth as arbitrary rather than the product of a process of reflective, shared public reason, and therefore irrelevant for justice as fairness.

Similarly, Rawls regards individuals' native endowments as arbitrarily acquired and therefore thinks that the distribution of such native endowments should play no role in the construction of principles of justice. Because one is not responsible for those features of oneself that one acquires at birth – such as race, natural abilities, social class, etc. – one should neither benefit from, nor suffer as a result of, those things.

Moreover, and consistent with his critique of utilitarianism, Rawls holds that individuals' pre-contractual determinate preferences and desires should not affect the content of principles of justice. He follows Rousseau in holding that determinate preferences and interests are the outcome of individuals' socialization under political and social institutions. Preferences and motives which were formed under less than just

⁴⁵ See Locke (1980).

institutions, therefore, should be subject to critical evaluation rather than automatically accorded normative weight in determining what the content of principles of justice should be.⁴⁶ For that matter, even those preferences and motives which were formed under just institutions should have to face critical scrutiny to ensure that they are acceptable from a moral point of view. If unexamined preferences and motives were given normative weight, then principles of justice would simply reflect and reinforce those existing preferences and motives, as well as existing inequalities in property and power. Justice, in that case, would be “nothing other than the advantage of the stronger” and would conflict with our common sense moral commitment to fairness.⁴⁷

To insulate the construction of principles of justice from objectionable motives and inequalities of property and power, then, Rawls offers a formal argument which relies on the idea that principles of justice are to be determined by “the choice which rational men would make in [a] *hypothetical situation of equal liberty*” (TJ: 11, my emphasis). In an original position we construct a “situation of equal liberty” by requiring that the parties who are asked to come to an agreement on principles of justice do so from behind a veil of ignorance. The veil of ignorance prohibits knowledge of “morally irrelevant” features of human beings so that the parties in the original position are unable to tailor principles to their own advantage. Rawls stipulates that

No one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not

⁴⁶ The interesting flip-side of this idea, as we’ll see when we examine Rawls’s approach to the fact of reasonable pluralism in chapter 5, is that comprehensive doctrines that would emerge under *just* institutions do require our attention as a matter of justice.

⁴⁷ See Thrasyarchus’s definition of justice in Plato (1974: 338c). While Plato’s reply to Thrasyarchus throughout the Republic relies on an argument about the Form of justice as an objective, metaphysical standard, Rawls’s own “political, not metaphysical” reply would rely on a Hegelian appeal to the, albeit inarticulate, principles implicit in the political culture of a constitutional democracy. More on this below.

know their conceptions of the good or their special psychological propensities (TJ: 11).

So constructed, the original position is “the *appropriate* initial status quo, and thus the fundamental agreements reached in it are fair. This explains the propriety of the name ‘justice as fairness’: it conveys the idea that the principles of justice are agreed to in an initial situation that is fair” (TJ: 11).

Thus, while Rawls says that “conceptions of justice must be justified by the conditions of our life as we know it or not at all” (TJ: 398), he does not hold that existing interests, motives, or the given distribution of resources and power should have normative significance. A conception of justice should acknowledge and wrestle with the problems of motivation and compliance, but in doing so we need not resign ourselves to an unattractive social and political status quo. Our common sense morality, if only we would reflect on it, would not allow us to resign ourselves to anything unfair. “The limits of the possible are not given by the actual, for we can to a greater or lesser extent change political and social institutions, and much else” (JF: 5). To be sure, “[i]f one starts directly with the particular case as known, and if one accepts as given and definitive the preferences and relative positions of the parties, whatever they are, it is impossible to give an analysis of the moral concept of fairness” (CP: 207n.13). Justice as fairness, then, uses a veil of ignorance in the original position to model a fair, albeit hypothetical, situation of initial equality as the appropriate status quo from which rational actors can choose defensible and legitimate principles of justice.

The veil of ignorance and the “hypothetical situation of equal liberty” that it offers, then, is a central feature of Rawls’s *hypothetical constructivist* approach to justice and the demands of realism and provides the centerpiece of a general strategy that could be used to deal with demands of realism. Using that general strategy, we ask agents to set aside their

determinate preferences, motives, and beliefs, and consider what principles they could agree to knowing only that they are rational and reasonable agents who will want to ensure fair conditions and sufficient resources to pursue *a* conception of the good, to participate in public political life, and to contribute to the development of their two moral powers (JF: 18-19). Because the agents who would reach an agreement in such a situation are not fully self-aware, the agreement must be regarded as *hypothetical*. But the conditions which permit us to speak merely of a *hypothetical* agreement are those same conditions which ensure that the content of that agreement – the conception of justice – is the outcome of reflection on relevant reasons, rather than a reconstruction of an unequal and unattractive social and political status quo. If features of the status quo or other demands of realism are to have an impact on the design of principles of justice, then it must be the case that the agents behind the veil of ignorance recognize good reasons for accommodating those demands. So are there any demands of realism that could pass the *hypothetical constructivist* test established by Rawls's theory?

II. Strains of Commitment and Unattractive Concessions?

Rawls's approach to the justice and motivation tension, and the tension between justice and the demands of realism more generally, depends crucially on a distinction between two stages of theory. At the first stage, principles of justice are selected by parties in an initial situation of equality and behind a veil of ignorance. That ensures, in Rawls's view, that principles are articulated and justified with attention concentrated on the attractiveness and fairness of those principles. At the second stage, the conception articulated at the first stage is tested against knowledge about citizens' interests, motives, and capacities to see if it would

be stable or motivationally possible. A defensible conception of justice, Rawls thinks, is one that could pass the tests of both stages.

But while Rawls states this as the official approach of justice as fairness, the distinction between the two stages seems blurred when concerns about the “strains of commitment” are raised in the first stage of principle construction. He writes that the parties in the original position “cannot enter into agreements that may have consequences they cannot accept” (*TJ*: 153). Thus, the parties in the original position “must ask themselves whether those they represent can reasonably be expected to honor the principles agreed to in the manner required by the idea of an agreement” (*JF*: 103)⁴⁸ Knowing that “the original agreement is final and made in perpetuity” contractors are to ask themselves whether they would be able to abide by and support the principles that are selected once they find themselves in a society guided by those principles. Indeed, “there is no second chance,” Rawls writes; “[a] person is choosing once and for all the standards which are to govern his life prospects” (*TJ*: 153). If the agreement that parties in the original position make has this condition of finality, then the parties must ask whether they could continue to accept its consequences once the veil is lifted and they learn what position they will hold in the society guided by justice as fairness. “[W]hen we enter an agreement,” Rawls writes, “we must be able to honor it even should the worst possibilities prove to be the case. Otherwise we have not acted in good faith” (*TJ*: 153).

On a weak reading of the strains of commitment, parties in the original position must ask themselves whether they could abide by the principles of justice regardless of the position in which they find themselves once the veil is lifted. That reading regards the strains of commitment test as a question of the rational grounds for compliance. Parties

would ask themselves whether they think that they could comply with principles of justice even in cases where those principles appear to constrain the pursuit of their self-interest. But Rawls wants more than that weak reading of the strains of commitment. He maintains that the parties in the original position are to agree to a conception only if they think that “when the conception is realized in basic institutions, those who grow up and live under them acquire a sufficiently strong sense of justice” (*JF*: 103n.26). On this strong interpretation of the strains of commitment, the parties must ask whether they could be motivated to support, and not simply grudgingly accept, the conception of justice. In that way, the stability of a well-ordered society could be ensured and that would give political agents another reason for agreeing to those principles to regulate the basic institutions of society.

However, it is possible that those principles that could pass the strains of commitment test might not be attractive principles from the point of view of justice. But Rawls himself says that “[j]ustice is the first virtue of social institutions....[L]aws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust” (*TJ*: 3). Asking the parties to consider whether they find the strains of commitment excessive certainly invites individuals to withhold their consent until concessions are made to their particular, and perhaps unattractive, motivational concerns. By giving a central role to requirements of democratic legitimacy and agreement in justice as fairness Rawls seems, on this objection, to open the door to unattractive concessions to citizens’ motivational weaknesses and perhaps also to the unequal distribution of resources and power in existing political societies.

⁴⁸ See also *TJ* sec. 29; *JF*: 103-4, 128-30; *CP*: 250-2; and *PL*: 17ff., as well as Rawls’s thoughts about stability and the sense of justice generally.

The strains of commitment certainly tug hard on potential contractors when they are permitted information about their particular situations. Agreement in the original position on Rawls's difference principle, for example, would be hard to reach when both the strains of commitment and particular information are part of the initial situation. Wealthy individuals who would face high rates of taxation of their current holdings if the difference principle were selected would have an incentive to withhold agreement until the principle of distribution conceded ground to their interests. Indeed, it is hard to see why any wealthy individual would agree to the difference principle when the original position is regarded as a place where rational individuals should aim to advance their self-interests. In the case of societies with many individuals who have developed a preference for positional goods or relative standing, for example, an agreement which satisfies the strains of commitment, if one could be reached, would look very different than Rawls's justice as fairness and I expect would be less egalitarian.⁴⁹ Why, then, does Rawls think that not only is it acceptable to

⁴⁹ It is useful to note how Rawls thinks that a competing conception of justice – namely, utilitarianism – would fare when faced with the strains of commitment. Parties in the original position, Rawls thinks, would regard the strains of commitment of utilitarianism as excessive and would therefore reject it. A utilitarian conception, in effect, asks that,

even when we are less fortunate, we are to accept the greater advantages of others as a sufficient reason for lower expectations over the whole course of our life. This is surely an extreme demand. In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems incredible that some citizens should be expected, on the basis of political principles, to accept lower still prospects of life for the sake of others" (*TJ*: 155).

Utilitarians, Rawls says, have to rely on the development of external moral virtues like benevolence and sympathy to ensure the stability of utilitarian justice rather than locating stability within the attractiveness of the conception of justice itself. Justice as fairness, by contrast, tries to locate motives and reasons for endorsing the conception that are internal to the conception itself. Justice as fairness, with its two principles of justice, aims to satisfy the strains of commitment by ensuring that "the least advantaged feel that they are a part of political society, and view the public culture with its ideals and principles as of significance to themselves" (*JF*: 129). Indeed, Rawls thinks that an attractive and stable conception of justice must satisfy the requirements of a strong interpretation of the strains of commitment which asks that the least advantaged under a given conception of justice actively endorse the conception rather than a weak interpretation which asks only that the least advantaged are satisfied enough that they don't "reject society's conception of justice and are ready to resort to violence to improve their condition" (*JF*: 129).

consider the strains of commitment, but that his conception of justice – justice as fairness – would survive that test?

In the first place, the veil of ignorance restricts the information that parties in the original have about themselves and therefore eliminates attempts to secure positional goods in the principles of justice selected. Rawls's contract is not the product of bargaining among individuals fully aware of their determinate preferences and motives. Samuel Freeman explains that the combination of the veil of ignorance with the strains of commitment entails that the parties in the original position will “strongly favor [Rawls’s] principles of justice over the principles of utility and other consequentialist views. For it is much more difficult for those who end up worse off in a utilitarian society to willingly accept their situation. Given what we know about human nature, the person is rare who can freely and without resentment sacrifice his or her life prospects so that those who are better off can have even greater comforts, honors, and enjoyments” (2003: 20). Because Rawls prohibits knowledge of particular information about the contracting parties in the original position attending to the strains of commitment does not give rise to certain individuals making unreasonable demands in return for commitment to the conception of justice. Behind the veil of ignorance, parties simply don’t know what determinate preferences and motives they have, so when they are asked if the strains of commitment are excessive, they have no reason to withhold their agreement in an effort to elicit concessions to determinate preference and motives. The parties in the original position, Rawls says, “have no basis for bargaining in the usual sense” (*TJ*: 120).⁵⁰

⁵⁰ Rawls acknowledges that problems of envy and preferences for positional goods do exist and that they “must be reckoned with” (*TJ*: 465). To reckon with them he says that “[a]t first we reason as if there is no problem of envy and the special psychologies; and then having ascertained which principles would be settled upon, we check to see whether just institutions so defined are likely to

Yet, while prohibition on particular knowledge in the original position prevents strategic bargaining, perhaps allowing knowledge of even the “general facts of moral psychology” could lead the parties in the original position to consider the problems of motivation in the wrong way. Parties in the original position might reason, for example, that it is “desirable to modify normative principles in response to anticipated non-compliance among others.”⁵¹ If we think that others might violate our initial, more demanding, principles then we might think it reasonable to make those principles less demanding in order to ensure compliance. That is, we are tempted to adjust the conception of justice not because we think citizens *cannot* comply with the more demanding principles but simply because we expect that some citizens *will not* comply. But if we are talking about modifying principles to account for anticipated motivational deficits then the principles we end up with would seem to be the *most feasible* principles of justice all things considered rather than ideal principles of justice. To put it another way, if the thought is that we should modify principles for practical reasons then we must already have an antecedent idea of what the most desirable principles are. And if we modify those antecedent principles to account for motivational concerns, or other demands of realism such as anticipated non-compliance, then we seem to have moved away from questions of normative justification and towards questions of practical application. Important questions, to be sure, but questions of application presuppose the existence of principles that can be applied.

There is the risk, then, that the content of the principles of justice chosen in the original position will accommodate unattractive motivational or compliance concerns, even

arouse and encourage these propensities to such an extent that the social system becomes unworkable” (*TJ*: 465).

⁵¹ Joe Heath stated the concern this way to me in an email. My understanding of the compliance and assurance problems has benefited from reading Elster (1989), Farrelly (2004), and Heath (2001, 2002).

while they avoid concessions to particular preferences or interests. So considering the strains of commitment at the first stage of theory construction, then, seems to put Rawls's justice of fairness at risk of making concessions for reasons of mere practicality that otherwise would have been avoided had the two stages of theory been clearly distinguished.

III. G.A. Cohen on Incentives

The work of G.A. Cohen on justice and incentives provides a concrete way to work through and to sharpen the problem of motivation and the strains of commitment.⁵² Cohen argues that because Rawls restricts the application of the difference principle to the basic structure of society rather than applying it also to the social ethos of citizens in the society, he permits individuals to make choices that violate the spirit of the difference principle itself. Indeed, Cohen maintains that not only does Rawls *permit* incentive-seeking behaviour, but that Rawls must regard that behaviour as not unjust. If correct, Cohen's claim means that justice as fairness entails an unattractive concession to self-interested motives. The talented agents who demand incentives, Cohen alleges, do so not because they are necessary, but simply because they want them and have the power to demand them successfully.

The difference principle holds that "social and economic inequalities are to be arranged so that they are...reasonably expected to be to everyone's advantage" (TJ: 53). But Cohen thinks that Rawls's defense of permissible inequality is not a normative defense, but instead "a merely factual defense of it" (2000: 120). "Despite what Rawls himself says," Cohen charges, "he does not show that incentive-based inequality is just, on his own conception of justice, but, at most, that it is regrettably unavoidable" (2000: 120-1). On

⁵² While in his work on incentives and his critique of Rawls, Cohen does not discuss the "strains of commitment" in particular, the concern that he articulates arises because of Rawls's consideration of the strains. See Cohen (2000).

Cohen's view, the incentive rewards that are "necessary" for the talented to be more productive (and thereby benefit the worse off) are necessary, if they are, because of choices made by the talented themselves. In Cohen's words:

Talented people who affirm the difference principle would find those questions [about the necessity of incentives] hard to handle. For they could not claim, *in self-justification*, at the bar of the difference principle, that their high rewards are necessary to enhance the position of the worst off, since, in the standard case, it is they themselves who *make* those rewards necessary, through their own unwillingness to work for ordinary rewards as productively as they do for exceptionally high ones, an unwillingness which ensures that the untalented get less than they otherwise would. High rewards are, therefore, necessary only because the choices of talented people are not appropriately informed by the difference principle (2000: 127).

Incentives and the distributional inequalities they introduce are not logically or physically necessary, but instead are the product of choice and power. In that case, Cohen thinks there is a discrepancy between Rawls's egalitarian intuitions and his (less than) egalitarian principles. Where incentives are necessary – that is, in the case where it appears that the difference principle has not become a part of the social ethos of the talented – Cohen thinks that the social ethos should be transformed to fit with the requirements of the difference principle.

From Cohen's perspective it seems that Rawls has made a bad deal with the talented in order to elicit their agreement. That is, Cohen accuses Rawls of watering down the requirements of justice, of moving away from a more attractive egalitarian view, in order to gain the adherence of those who would prefer to retain some of the benefits they might receive from the use of their talents in a system that appears to reward talents – talents which, Rawls himself admits, are arbitrary from a moral point of view and should be viewed as a "common asset". The difference principle – or, more precisely, the scope of inequality which Rawls thinks is permitted by the difference principle – is the product not of reflection on the requirements of justice, but instead on the requirements of stability under non-ideal

conditions. On this view, Rawls seems to hold that the motivational weakness of the talented is reason enough to adjust the principles of justice and that the parties in the original position would agree when they are asked to consider the strains of commitment.

Rawls thinks that he has “reconcile[d] us to our political and social condition” (*LP*: 11). But, like Rawls, Cohen thinks that a conception of justice must not simply accommodate the existing status quo, otherwise it loses its critical and emancipatory appeal. Recall again that Rawls holds that “[j]ustice is the first virtue of social institutions” and that “laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust” (*TJ*: 3). If Cohen is right that in the very content of the principles of justice Rawls makes concessions to motivational deficits which are both unattractive and unnecessary, then Rawls has failed on his own terms. To rescue Rawls’s position from Cohen’s objection, then, we need to show either that Rawls does not permit the sort of inequalities that Cohen accuses him of permitting or, if he does permit such inequalities, that permitting them isn’t, as Cohen thinks, a matter of accommodating unattractive and unnecessary motivational deficits but instead a matter of accommodating attractive motivational concerns or the limits of human motivational capacity.

The question that parties in the original position face might be stated this way: Would the strains of commitment be excessive if, rather than requiring the difference principle to apply to the basic structure of society, justice demanded that it apply also to the choices political agents make? To Cohen the question is unimportant because justice should not depend on agents’ agreement. That can only open the door to concessions to unattractive preferences and interests. But for Rawls, because justice is a matter not simply of equal distributions, but of what best serves the fundamental interests of human beings concerned

with the development of their two moral powers, some motivational concerns make that question, and consideration of the strains of commitment, necessary.

Basic Structure, Social Ethos, and Justice

Before proceeding to a direct evaluation of Cohen's critique of Rawls in the next section, I want to offer a few words to clarify Rawls's view about the relationship between the basic structure, the social ethos, and justice. Cohen seems to think that Rawls's claim that justice is a subject for the basic structure of society entails that endorsement of justice as fairness is not required at the level of the social ethos. While Cohen is right to maintain that justice as fairness does not require individuals to make distributional calculations and decisions at the level of their individual, daily behaviour – and thus that they are not prohibited from making incentive demands – it is not true to say that justice as fairness does not have a place in the social ethos at all. In Rawls's view, justice requires that the difference principle apply to social and political institutions. And that means that, at certain points in time, we should assess whether or not the social and political institutional configuration is producing distributional results consistent with the difference principle and that we should alter the institutions if necessary.⁵³

But contrary to Cohen's assumption, the fact that the difference principle *applies* to the basic structure does not mean that citizens themselves are released of any obligation to *endorse* the difference principle and to assess, reform, and support an institutional configuration which satisfies the difference principle. The difference principle and the demands of justice as fairness more generally should be endorsed at the level of the social ethos on Rawls's view. But that means only that citizens have an obligation to support and

to assist in the construction of social and political institutions which would satisfy the difference principle. Citizens do not have, on Rawls's view, a duty to comply with the distributional demands of the difference principle at the level of individual distributional decisions.⁵⁴ *Endorsement* of the difference principle at the level of the social ethos – which justice requires – does not entail a change in the *content* of the difference principle.

Endorsement of the difference principle by citizens, then, requires only that they really believe and ensure that the institutional configuration produces inequality only if it is to the advantage of the least advantaged.

Of course, this does not mean that there is not still an important challenge in Cohen's critique. Indeed, Cohen can concede the point that Rawls's view requires *endorsement* of the difference principle at the level of the social ethos while still holding that the *content* of the difference principle should be revised to apply to the behaviour and decision-making of individuals. So can a convincing reply be offered to Cohen's concern?

IV. Replies to G.A. Cohen

In this and the following sections I will reply to Cohen's concern by emphasizing the roles that democratic legitimacy, primary goods and the conception of the person play in Rawls's

⁵³ For Rawls's account of the social and political institutions associated with distributive justice, see *TJ*: § 43 "Background Institutions for Distributive Justice".

⁵⁴ Josh Cohen, who has helped me to see this distinction, has provided an insightful example: Supporting the First Amendment's prohibition on laws regulating free speech – i.e., that *Congress* shall pass no law which abridges the freedom of speech – is something that citizens can support at the level of the social ethos. But supporting the First Amendment which requires that *Congress* pass no law abridging freedom of speech does not require individual citizens to consider whether their own actions constitute restrictions on free speech. In other words, if people endorse and really believe in the First Amendment that means only that they really believe that it should apply to *Congress* and its law-making behaviour. Similarly, then, if people really believe in and support the difference principle, that means only that they really believe that it should apply to the basic structure. Justice is a matter of the basic structure which must be endorsed at the level of the social ethos, but endorsement at the level of the social ethos does not alter the content of the demands of justice in the way that Cohen seems to think it does.

justice as fairness. Unlike other replies to Cohen's critique, my reply maintains that Rawls's restriction of the principles of justice to the basic structure is a feature of justice as fairness that reasonable political agents would have chosen, not to accommodate some unattractive and unnecessary motivational deficit, but rather to ensure that the conception of justice is one that individuals could regard as promoting their fundamental interests as free and equal citizens.⁵⁵

Other replies to Cohen rely on a number of strategies. Joshua Cohen's reply, for instance, is instructive in the way he shows how, contrary to G.A. Cohen's claims, justice as fairness is concerned with the social ethos of a well-ordered society.⁵⁶ In the first place, J.Cohen (2002) notes how justice as fairness does not simply take the existing, perhaps inequalitarian and egoistic, ethos as fixed in the way that G.A. Cohen suggests. The difference principle permits inequalities, if it does at all, not on the basis of determinate or particular preferences or motives that some individuals might have. Rather, inequalities are permitted only on the basis of the general facts of human psychology.

As I have noted, however, Rawls fails to clearly specify the content of the so-called general facts of human psychology. It might be the case that egoistic motives are a feature of that general psychology. Thus, G.A. Cohen can respond that the veil of ignorance on particular information is insufficient to exclude this objectionable feature of human

⁵⁵ Notable papers which wrestle with Cohen's incentives, structure, and ethos critique of Rawls's theory (as we might call it) include Williams (1998), Estlund (1998), Murphy (1999), Pogge (2000), Cohen (2002). Wolff (1998) also takes up the issue of ethos versus institutions but is less a direct reply to Cohen's critique of Rawls than the other works cited here. While the essays by Williams (1998) and Pogge (2000) give some mention of autonomy and democratic legitimacy, neither of them match the degree to which Murphy (1999) and J. Cohen (2002) consider these concerns. And even these latter two don't capture the force of the autonomy and democratic legitimacy concerns that I emphasize here.

⁵⁶ In this paragraph I refer to G.A. Cohen and J.Cohen using the initials of their given names along with their shared surname in order to avoid confusion. Elsewhere in the chapter, where I simply write "Cohen", one should assume that I mean G.A. Cohen.

psychology from the original position. Even so, the force of J. Cohen's reply rests on the claim that Rawls's focus on the basic structure is not as objectionable as G.A. Cohen might think because "at least in some circumstances a Rawlsian conception of justice will require...changes in the social ethos because that is *how* institutional changes lead to a more just distribution" (2002: 364). J. Cohen emphasizes that Rawls's understanding of the effects that institutions have on a social ethos is inspired by Rousseau's statement "that a man is naturally good, and that it is solely by [our] institutions that men become wicked." If bad institutions make people wicked, then perhaps just institutions make them more just. Moreover, J. Cohen's reply captures the idea, discussed above, that at the level of the social ethos citizens might have an obligation to support and to work towards the construction of social and political institutions that satisfy the requirements of the difference principle even if they are not required to apply something like the difference principle to their individual decision-making and daily behaviour. Contrary to G.A. Cohen's claim that Rawls ignores the social ethos, then, J. Cohen shows how Rawls's concern for the social ethos is demonstrated by his view about how the social ethos can be transformed and what citizens are obligated to do to support justice.

Two other replies to Cohen are of particular interest in that they show how accommodations of human motives can be made for unobjectionable reasons. David Estlund (1998), for example, makes the case that motives of affection "might compete with the demands of social justice" but not for egoistic reasons. Behaviour motivated by affection for one's spouse or children, for example, may contribute to inequalities in the wider society, but the motive itself can be regarded as "untainted" (1998: 102, 108). Moreover, attempts to replace the motives of familial affection with a society-wide "pseudoaffection" may have worse consequences for an egalitarian ethos than would

maintaining the status quo motives of affection. Thomas Pogge (2000) also acknowledges certain motives that might result in greater inequalities but are unobjectionable from a moral point of view. He notes, for example, the “fact that people have diverse preferences over jobs” which can motivate individuals to demand higher wages, not as incentives to greater productivity, but to compensate them, in some way, for agreeing to perform more productive work that they would sincerely prefer not to do.⁵⁷

While both Estlund and Pogge demonstrate that attention to motives, and therefore not all inequalities, are objectionable, that doesn’t suffice to show why incentives and therefore the inequalities that result, are necessary for ensuring the greatest benefit to the least advantaged. The difference principle holds that “social and economic inequalities are to be arranged so that they are...to the greatest benefit of the least advantaged” (*TJ*: 266) which suggests that inequalities are permissible when are *necessary* to improve the advantage of the least advantaged, not simply that inequalities which arise from unobjectionable motives are permissible. It may be permissible to accommodate certain untainted motives, but mere permissibility does not demonstrate necessity. And it is a demonstration of *necessity* that the difference principles demands. Thus, if Rawls’s conception of justice is the product of a hypothetical agreement, then we need not only reasons why the parties in the original position might regard accommodations of untainted motives as unobjectionable, we need to show also that there are reasons why the parties would think agreeing to these concessions would be good – that is, required by justice and not merely a practical concession. The parties in the original position should, and Rawls thinks that they will, agree only to accommodate those motives and only those features of human psychology that promote some interest that they themselves have. But what interest or interests do parties have when

⁵⁷ As Joshua Cohen has pointed out to me, Pogge’s shift from the idea of “incentives” to the idea of

they are in the original position which could guide their judgments between merely permissible and necessary concessions to certain general facts of human psychology?

V. From Facts to Conceptions

Note that in restricting the difference principle to the basic structure, Rawls is not responding to the demands of citizens' determinate preferences and motives. With the veil of ignorance in place, no one would be in a position to know whether concessions to incentive-seeking or egoistic behaviour would benefit or harm them once the veil is lifted. So Rawls's conception does not accommodate the demands of any particular class or individual when it restricts application of the difference principle to the basic structure. But our question was whether Rawls's conception of justice makes an objectionable concession insofar as it relies on the general facts of human psychology. Indeed, Rawls himself writes that "the theory of justice assumes a definite limit on the strength of social and altruistic motivation. It supposes that individuals and groups put forward competing claims, and while they are willing to act justly, they are not prepared to abandon their interests" (*TJ*: 248). So we might wonder whether that might lead to concessions to objectionable motives.

Even there, however, Rawls does not make concessions to the empirical facts of human psychology, but instead tries to accommodate a *conception* of the person. That conception "is meant as both normative and political, not metaphysical or psychological" (*JF*: 19).⁵⁸ It holds that citizens are to be regarded as free and equal, and that citizens have two moral powers – a capacity for a sense of justice and a capacity for a conception of the

"compensation" actually changes the issue rather than addressing it directly.

⁵⁸ In *Justice as Fairness* Rawls writes that "[t]his conception of the person is not to be mistaken for the conception of a human being (a member of the species *homo sapiens*) as the latter might be specified in biology or psychology without the use of normative concepts of various kinds, including, for example, the concepts of the moral powers and of the moral and political virtues" (*JF*: 24).

good (*JF*: 18-24). Notice how that conception expresses the Rousseauian theme of Rawls's political theory: Citizens are regarded as having a fundamental interest in having the freedom to pursue a conception of the good and the aim of a conception of justice is not only to acknowledge that interest, but to secure conditions necessary for the protection and encouragement of that freedom. What the representatives in the original position are to ask, then, is what they would need in order to ensure that all are regarded as free and equal and to ensure that they have the resources necessary to exercise the two moral powers.

Framing the question of justice in terms of the requirements of a normative conception of the person leads Rawls to re-characterize his idea of primary goods. In the Preface to the French Edition of *Theory*, Rawls says that “a serious weakness of the original English edition was its account of primary goods” (*CP*: 417). In the earlier account, primary goods were held to be those things that rational persons would want whatever else they might want because they were characterized as those material and social resources that individuals would need in order to pursue their aims and ends (whatever those might be). But in the years after the publication of *Theory*, Rawls began to see how principles of justice could not be articulated by asking simply what rational actors would choose under conditions of limited information.⁵⁹ Instead, he says that “persons are to be viewed as having two moral powers...and as having higher-order interests in developing and exercising those powers. Primary goods are now characterized as what persons need in their status as free and equal citizens, and as normal and fully cooperating members of society over a complete life” (*CP*: 417). Primary goods in the later characterization are to be seen as “answering to their needs as citizens as opposed to their preferences and desires” (*CP*: 417).

⁵⁹ See *JF*: 82n2, and *CP*: 365n5.

What G.A. Cohen regards as Rawls's accommodation of egoism should be regarded instead, in light of the conception of the person, as an attempt to accommodate certain fundamental interests of citizens. Whereas G.A. Cohen thinks that distributive equality is the primary subject of justice, Rawls's maintains that a concern for equality must be accompanied by a concern for citizens' interest in freedom, their autonomy as citizens. The conception of the person holds that individuals have a capacity to have and to revise a conception of the good and justice as fairness maintains that citizens' having and pursuing those conceptions should guide our construction of principles of justice. If freedom entails that people "regard themselves as being entitled to make claims on their institutions so as to advance their conceptions of the good" (*JF*: 23), then principles of justice which are not responsive to citizens' understandings of their reasonable conceptions fail to recognize a fundamental human interest in freedom. Of course Rawls recognizes that a conception of justice cannot deliver a "society in which all can achieve their complete good, or in which there are no conflicting demands and the wants of all fit together without coercion into a harmonious plan of activity" (*TJ*: 249). That would be a society, as Rawls says, "in a certain sense beyond justice" – a society which "has eliminated the occasions when the appeal to the principles of right and justice is necessary" (*TJ*: 249). But if such a utopia is unattainable, then justice requires at least an effort to attend to and to accommodate in some way citizens' interest in having the freedom necessary to pursue a reasonable conception of the good even if cannot guarantee universal success is all those pursuits.

By accommodating the normative conception of the person, rather than mere empirical facts about human beings, Rawls's conception not only avoids Cohen's critique, but indicates one way in which a conception of justice should be responsive to citizens' motives and the demands of realism more generally. Indeed, a conception of justice which

failed to acknowledge and in some way accommodate citizens' interest in freedom would not only be unstable – would not only create excessive strains of commitment – it would be an unattractive conception. Meeting the requirements of the strains of commitment, Rawls thinks, solves part of the problem of motivational possibility. That is, a conception of justice is endorsed by the parties to the original agreement only if they think that under that conception all citizens could, over time, develop an appropriate sense of justice and willingly, rather than grudgingly, support the conception. But, because the strains of commitment question is framed in a subjective rather than an objective way – that is, because it asks the parties in the original position to ask themselves whether they think the strains of commitment are acceptable or excessive – that requirement attends not only to the problem of stability, but in some ways also to that of legitimacy. Justice as fairness demands that a sense of justice which supports the two principles of justice develops *in virtue of* the conception of justice it defends. Justice as fairness aims for stability *through* legitimacy.

However, as far as the difference principle itself is concerned, G.A. Cohen's objection seems to hold. The replies which point to "untainted" motives or, in my case, to the needs of democratic legitimacy, do not demonstrate why inequalities would be *necessary to improve the condition of the least advantaged*. Making justice a matter of the basic structure rather than the social ethos and thereby creating some space for the exercise of liberty and some associated distributional inequality might be necessary from the point of view of justice as fairness as a whole – that is, a point of view which is concerned to reconcile liberty and equality – even if it is not ideal from the point of view of the least advantaged once the veil of ignorance is lifted. From the perspective of justice as fairness as a whole, such inequalities may be permissible insofar as they are the outcome of how the conception of justice responds to freedom-based motives to which it must respond if the conception is to be justifiable and

legitimate. Moreover, given that the parties in the original position are assumed to be concerned with material resources not as mere rational economic agents but instead as democratic citizens, it might be reasonable for them to agree to principles which permit some material inequality which isn't strictly necessary to improve the advantage of the least advantaged. And insofar as the parties in the original position must take up all of these different points of view when deliberating about justice and the strains of commitment, it is reasonable to hold that justice as a whole does not require satisfying what would be considered ideal from only one of the many points of view that the parties must take on.⁶⁰

VI. Conclusion

G.A. Cohen's criticism of Rawls's justice as fairness, it seems, derives at least in part from the assumption that justice is primarily a matter of distributive justice. But as I have tried to show, Rawls is concerned not simply with distributive justice, but with justice in a larger sense – with justice understood as fair terms of social cooperation. Citizens have interests not simply in material resources, Rawls holds, but in how a package of rights, liberties, and resources together will affect their ability to participate in the political life of their societies as equal democratic citizens and their liberty to pursue a conception of the good. A defensible theory of justice, in other words, tries to articulate fair terms of social cooperation in a way which recognizes that citizens have an interest in arrangements which permit the pursuit of both public and private aims. Once that idea enters the picture, a conception of justice must accommodate itself to at least some of the motivational concerns of citizens – that is, to one of the demands of realism.

⁶⁰ I am indebted to Sarah Song for help in clarifying the argument of this paragraph.

Of course, G.A. Cohen and others in the utopian camp might say that justice is not *for* human beings but instead an independent idea that does not or should not depend on human motives or other demands of realism in any sense. But then it would be hard to see why we should deliberate and care about justice at all. It might be an object that we study out of curiosity, but its connection to the political and social lives of real people would be lost. And yet, even if we regard justice as something which is *for* human beings rather than an independent object of curiosity, Cohen's worries about the way concerns about realism, legitimacy, and stability might affect the content of a conception of justice are instructive. They show us that a distinction between relevant and irrelevant demands of realism must be made and, moreover, that concessions to objectionable features of the social and political status quo should be avoided. The Rawlsian reply to Cohen's worries here have shown, I think, how we can at least begin to make that distinction and how we can avoid concessions to objectionable demands of realism in the selection of principles of justice.

However, there are two outstanding issues that we will have to resolve over the course of the next two chapters: In the first place, it is still not clear why legitimacy and agreement need to play such a fundamental role in the construction of a conception of justice. Perhaps the concerns about realism can be handled without the idea of agreement. The next chapter considers the way in which Habermas's political theory offers an answer to that concern and, moreover, reconsiders Rawls's political theory in light of the more democratic approach to the problem of motivation that Habermas appears to defend. The other outstanding issue has to do with the way the Rawlsian approach to the demands of realism fares in light of the diversity of moral and religious worldviews which characterizes contemporary democratic societies. A hypothetical constructivist approach to the demands of realism might falter under conditions of pluralism and thus Rawls's political theory might require further

revision. I take up this issue in chapter 5. Our final assessment of Rawls's approach to the demands of realism, then, will have to be postponed until these outstanding issues can be addressed.

4

Habermas: Real Proceduralism

Tossed to and fro between facticity and validity, political theory and legal theory today are disintegrating into camps that hardly have anything to say to one another. The tension between normative approaches, which are constantly in danger of losing contact with social reality, and objectivist approaches, which screen out all normative aspects, can be taken as a caveat against fixating on one disciplinary point of view.

- Jürgen Habermas (BFN: 6)⁶¹

Interests can be satisfied through generalized behavioral expectations in the long run only if interests are connected with ideas that justify normative validity claims; ideas can in turn gain broad empirical acceptance only if they are connected with interests that lend them motivational force.

- Jürgen Habermas (BFN: 69)

Jürgen Habermas's political theory is guided by the idea of a "self-organizing community of free and equal citizens" (BFN: 445, 7). He defends a discursive conception of democracy which tries to embrace Rousseau's conviction, on the one hand, that "human beings act as free subjects only insofar as they obey just those laws they give themselves" (BFN: 445-6) and Rawls's claim, on the other hand, that principles of justice should not be tainted by objectionable interests nor influenced by power inequalities. At the same time, however, Habermas departs from Rousseau's *monological* approach – the idea that the moral point of

⁶¹ Habermas's works are cited in the essay as follows: BFN: *Between Facts and Norms*, trans. William Rehg (Cambridge: MIT Press, 1996); DE: "Discourse Ethics: Notes on a Program of Philosophical Justification" and MCC.4: "Moral Consciousness and Communicative Action" both in *Moral Consciousness and Communicative Action* (Cambridge: MIT Press, 1990); RDE: "Remarks on Discourse Ethics" in *Justification and Application: Remarks on Discourse Ethics* (Cambridge: MIT Press, 1993); STPS: *The Structural Transformation of the Public Sphere*, trans. Thomas Burger (Cambridge: MIT Press, 1989); and RPUR: "Reconciliation Through the Public Use of Reason" in *The Inclusion of the Other* (Cambridge: MIT Press, 1998).

view is located in the “inner voice” of a reflective individual. And he rejects Rawls’s *hypothetical constructivism* – i.e., the attempt to insulate deliberation about principles of justice from the determinate preferences, motives, and beliefs of citizens in an original position. Instead, Habermas defends a *real proceduralist* variant of realistic utopianism which relies on principle of discourse ethics which holds that “only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse*” (MCCA: 66). He aims for a more democratic realistic utopia which gives citizens a central role in discourses about those norms which will apply to institutions and actions while it simultaneously filters out objectionable motives via an impartial point of view implicit in the pragmatic presuppositions of argumentation.

In this chapter, I investigate Habermas’s discourse ethics which ties the justification of principles of justice to actual discourses among citizens who are fully aware of their own preferences, motives, and beliefs. I ask whether the actual deliberation Habermas advocates can deliver on the promise that individuals can and will look beyond their own interests, preferences, and needs and consider the interests, preferences, and needs of others in testing appropriate moral norms. While Habermas thinks that impartiality is achieved not by abstracting from features of human beings and the social world, but instead by permitting a free exchange of information, opinions, and arguments among fully aware citizens, we might wonder whether those discourses have undesirable consequences for the moral norms that are endorsed by deliberating citizens. Indeed, given that political discourse in existing capitalist societies is permeated by inequalities of power and a variety of unattractive private preferences and expectations, it is hard to see at the outset how principles of justice that are the outcome of a *real proceduralism* embedded in an objectionable status quo could avoid the dangers of strategic bargaining and concessions to unfair advantages. They might offer a

more immediate solution to the problem of motivational possibility and capture some of the force of the democratic ideal of self-legislating citizens, but those gains might be achieved at the cost of principles which make concessions to objectionable motives as well as unattractive features of existing capitalist societies.

I argue that while Habermas's attempt to democratize the procedure for discovering appropriate norms is an important development for realistically utopian theory, his theory ultimately fails to provide an attractive approach to the justice and motivation tension. First, his failure to acknowledge the substantive requirements of ideal discourse exposes candidate principles to threats from objectionable motives and unattractive features of the social status quo. That failure, moreover, makes it hard to see how Habermas can defend a solution to the problem of motivational possibility which relies on citizens being able to develop a "postconventional stage of moral consciousness" (*DE*: 109) – that is, a psychological disposition to be motivated to act on principles because they are valid and not because they fit with any pre-political motives, preferences, or beliefs that the agent may have. Secondly, Habermas's move from a defense of actual discourses of justification to a defense of a "subjectless" flow of reasons in the formal and informal public spheres threatens the normative and motivational advances made by a discourse-theoretic conception of democracy (*BFN*: 299). In his attempt to "unburden" citizens of the cognitive, motivational, and organizational demands of complex societies by shifting from inclusive discourses to a more anonymous flow of reasons and the use of a legal system to sanction noncompliance, Habermas abandons the mechanism by which transformations of citizens' pre-political motives and preferences can be achieved and which can encourage the development of the "postconventional stage of moral consciousness".

In short, I argue that a concern for citizens' capacity to support valid principles of justice is unavoidable once we recognize that the legal and political institutions which Habermas thinks can sidestep the motivational problem ultimately depend on citizens' sense of justice and their designing and reforming institutions in light of that sense of justice. A theory that fails to demonstrate how that sense of justice can develop fails to offer an account of motivational possibility and, thus, fails to solve the problem of justice and motivation. In that sense, Habermas's *real proceduralist* approach – which regards justified principles as those which emerge from an actual process of democratic deliberation – fails to offer a compelling solution to the more general problem of how to reconcile justice to the demands of realism. Because the institutional configuration Habermas envisions fails to produce in citizens the capacities and commitments necessary to support defensible principle of justice, the *real proceduralist* theory that he defends fails to show how discourse participants will be able to distinguish between demands of realism which should, and those which should not, be accommodated by a conception of justice.

As in the previous chapter on Rawls's *hypothetical constructivism* this chapter on Habermas's *real proceduralism* will focus on concerns about motives and motivational possibility. The other demand of realism with which we are concerned – pluralism – will be addressed from time to time in this chapter, but a more comprehensive treatment of that concern will be postponed until Chapter 5. In this chapter, then, after providing a sketch of the main features of discourse ethics (Section I), I discuss the two ways in which Habermas's political theory fails to address the tension between motivation and principles of justice in a suitable manner. Section II focuses on the role of substantive justice in realizing the ideal speech situation which is central to Habermas's theory. Section III interrogates Habermas's move from direct to more impersonal discourses in *Between Facts and Norms* and his move to

rely on a legal system with sanctions for noncompliance as a strategy to unburden citizens of the problem of motivation altogether. I argue that these moves prevent discourse ethics from facilitating the transformation of citizens' preferences and motivations as necessary. Finally, in section IV, I offer some concluding reflections and point towards the outstanding problem of pluralism which I confront directly in the next chapter.

I. Radical Democracy and Discourse Ethics

Habermas's realistically utopian approach to the problem of motivation, and the demands of realism more generally, like those examined in earlier chapters, attempts to distinguish between objectionable and unobjectionable motives from a moral point of view and offers a solution to the problem of motivational possibility that is consistent with that point of view. Like Rawls, Habermas is guided by an ideal of a democratic society implicit in the public culture which includes a conception of the citizen as free and equal. But unlike Rawls, Habermas is not prepared to develop and defend substantive principles of justice in his theory because he thinks that working out the implications of the implicit democratic ideal for principles of justice is not a matter for philosophers alone but a matter for all citizens. Thus, he defends a procedural conception of democracy which outlines conditions under which principles could be accepted, and therefore justified, by actual citizens deliberating among themselves. As Thomas McCarthy describes the project, "rather than contractual arrangements among 'unencumbered' individuals with arbitrarily chosen ends, [Habermas's discourse ethics] involves processes of reflective argument among previously socialized subjects whose needs and interests are themselves open to discussion and transformation" (1990: xi). On this model the distinction between motives which should be accommodated in principles of justice and those which should be transformed or ignored is made by citizens

themselves albeit through the guidance provided by presupposed rules of argumentation which exclude certain claims and arguments.

While our focus in this chapter will be on Habermas's approach to the narrower concern about motives, it can be said that the way in which Habermas deals with that concern is illustrative of his approach to the demands of realism more generally. His *real proceduralist* variant of realistic utopianism holds that decisions about the demands of realism and whether they should affect the content of valid principles must be made by citizens themselves in deliberation with others. While Habermas is concerned, especially in his later work *Between Facts and Norms*, to "unburden" citizens of not only the motivational demands of right and justice, but also the cognitive and organizational complexities of contemporary societies (BFN: 114), and thus shifts to what we might call a more institutional or structural proceduralism, he nevertheless retains the idea that decisions about the demands of realism must be made in a deliberative fashion where participants have access to all information. To be sure, distinctions between information which is relevant or irrelevant are to be made by the deliberators themselves rather than by political philosophers in advance of actual deliberation. While the shift to an institutional or structural proceduralism is problematic, as I will argue later in the chapter, a real proceduralist approach to the tension between justice and the demands of realism is in many ways compelling because of the strong commitment it shows to the idea of legitimacy. Whereas *monological* realistic utopianism and *hypothetical constructivism* demonstrate a concern for agreement, a *real proceduralist* variant of realistic utopianism captures the idea that hypothetical agreements or agreements made in isolation from other citizens do not produce the same kind of legitimacy as agreements made by actually deliberating citizens.

Discourse Ethics and the Moral Point of View

In contrast to democratic theories which hold that democracy is a matter of uncritically aggregating given interests or selecting elites to rule until voters replace them with a less objectionable group, Habermas defends a much more rigorous conception which invites wide-spread participation and deliberation among all citizens about the fundamental principles which should guide their collective political life. Citizens are not simply left to discuss particular policies or regulations with an already established set of liberal rights limiting the agenda of that discussion. According to discourse ethics, “only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse*” (DE: 66). And it is in such practical discourses that Habermas locates the point of view from which distinctions between permissible and impermissible motives should be made. In this sense, the *moral* point of view that Habermas offers draws on the ideal of a democratic society but might be referred to more accurately as a *political* point of view because of the way in which Habermas locates normative validity in the outcome of the practical political discourses of citizens.

At the same time, Habermas has argued elsewhere that political discussion in existing capitalist societies is permeated by power inequalities and strategic maneuvering. Indeed, he rejects Hobbesian conceptions of justice because they appear to concede ground to objectionable motives as well as uncritically accept potentially unjust features of the social and political status quo. Contracting parties on Hobbes’s account, according to Habermas, “are equipped only with freedom of choice” and therefore “can justify their contractual agreements exclusively on purposive grounds, with the result that their reasons remain tied to the contingent interests and preferences of participants” (DE: 27). That tie to contingent interests and preferences is unacceptable because the interests and preferences of citizens in

modern capitalist societies are often illegitimately manipulated by private advertisers, corporate media, and state propaganda.⁶² Like Rousseau and Rawls, Habermas wants a conception of justice that does not amount to a mere bargain among individuals whose determinate motives are taken at face value. But why, then, does Habermas endorse a discursive approach to justifying moral norms which threatens to expose principles to objectionable motives and other unattractive preferences, strategic compromises, and power inequalities?

To answer these questions, it helps to examine what Habermas finds unacceptable in the Rawlsian (and Kantian) conception of justice. On Habermas's reading, Rawls "wants to ensure the impartial consideration of all affected interests by putting the moral judge into a fictitious 'original position'" where he, as much as any other individual so situated, "can undertake to justify basic norms on his own" (*DE*: 66). Because "the same holds true for the moral philosopher himself," Habermas claims that Rawls must view his principles of justice "not as the contribution of a participant in argumentation to a process of discursive will formation regarding the basic institutions of late capitalist society, but as the outcome of a 'theory of justice,' which he as an expert is qualified to construct" (*DE*: 66).⁶³ Habermas seems to agree with those critics of Rawls's veil of ignorance device who, on the one hand, hold that individuals are simply incapable of adequately abstracting from their own particular preferences or conception of the good, and those who, on the other hand, hold that

⁶² For Habermas's diagnosis of the ills of modern capitalist societies and analysis of the manipulations of social consciousness see the latter half of his *Structural Transformation of the Public Sphere*, the whole of *Legitimation Crisis* and, on the media, *Between Facts and Norms* (especially pp. 376-379).

⁶³ Rawls (1996) disagrees with this characterization of his view. See especially (1996: 382-5). More on this below in Section II.

individuals are incapable of adequately adopting the perspectives of others when considering the consequences of principles or actions on their interests and conceptions of the good.⁶⁴

The Benefits of Real Proceduralism

So how can arrangements which are more discursive or democratic rather than *monological* or *hypothetical* improve our ability to articulate principles of justice? In the first place, discursive approaches potentially lead to “a better-informed perspective” on the normative and practical concerns of principles of justice and public policies (Mendelberg, 2001: 3).⁶⁵

Whereas an individual deliberating alone knows her own interests and capacities and has limited access to information about the world and potential consequences of various principles and policies, a more discursive approach can produce information about interests, motives, and consequences that the solitary individual did or could not have.⁶⁶ Second, discursive democratic arrangements can help us to better develop and defend reasons which could justify principles of justice because of the opportunities they create for assessment of those justifications from multiple perspectives. Third, discursive approaches offer opportunities for individuals to become more aware of the interests and perspectives of others and perhaps improve the concern that individuals have for the consequences of decisions on others’ interests and well-being rather than simply their own.⁶⁷ Fourth, and relatedly, democratic approaches are potentially superior to monological and hypothetical approaches in that they reduce the possibility that deliberators will either misrepresent or

⁶⁴ See Deveau (2000), Sandel (1982), and the contributions to Benhabib & Cornell (1987), especially those by Benhabib and Young.

⁶⁵ See also Gutmann & Thompson (2004: 11-12).

⁶⁶ Of course, discursive approaches might also increase the quantity of *irrelevant* information and so a mechanism for sorting through the increased information would be necessary for the discursive approach to be informationally superior to monological deliberation.

⁶⁷ Gutmann & Thompson (2004: 10-11); Mendelberg (2001: 3, 4); *DE*: 65; and *RDE*: 48-9.

discount the interests and perspectives of others.⁶⁸ Finally, democratic or real proceduralist approaches have the potential to increase the legitimacy of, and therefore the motivation to support, principles which emerge from deliberation because participants can regard themselves as having contributed to their development rather than simply passively accepted them from some other source.⁶⁹

On Habermas's view, "nothing better prevents others from perspectively distorting one's own interests than actual participation" (DE: 67). Indeed, actual deliberation among citizens aware of their own interests ensures that the practice of taking the perspective of others is supervised and, if necessary, corrected by those very others whose perspectives are being taken. So "moral justifications are dependent on argumentation actually being carried out, *not for pragmatic reasons* of an equalization of power, but for internal reasons, namely that *real argument makes moral insight possible*" (DE: 57). That is not to say, however, that determinate motives are beyond criticism and possible transformation or rejection in moral discourse. Rather, determinate motives, as well as preferences, interests, and interpretations, are subject to critical examination and assessment by other participants when they "adopt the perspective of all others" (DE: 65).

More generally, an actual process of deliberation allows citizens to discuss the relevance of specific demands of realism as well as the reasons that permit or prohibit accommodations to any particular demand. A real proceduralist approach to the demands of realism has an advantage over monological or hypothetical approaches, then, insofar as it permits a fully informed discussion about potential demands of realism and allows each

⁶⁸ DE: 65, 67; RDE: 48-9, 58.

⁶⁹ I think that there are other benefits to deliberative democratic approaches, including and especially their encouragement of the development and use of citizens' capacity for critical reflection on their own interests, which I will discuss in chapter 6. My concern here, however, is not with the

citizen an opportunity to make a case for or against accommodation. Because monological and hypothetical approaches restrict the scope of participation by making decisions about practical and normative relevance in advance of actual deliberation, those approaches are at risk of overlooking important information or arguments that would be more likely to emerge where all citizens are allowed to offer their views at all stages of principle construction.

Moreover, “impartiality of judgment,” Habermas writes “is essentially dependent on whether the conflicting needs and interests of *all* participants are given their due and can be taken into consideration from the viewpoint of the participants themselves” (*RDE*: 48-9). “The descriptive terms in which each individual perceives his interests,” he maintains, “must be open to criticism by others” (*DE*: 67). Distinctions between legitimate and illegitimate motives and other demands of realism are not made *prior* to public discussion but *through* public discussion. Participants engage in a practice of ideal role-taking in which they consider the appropriateness of their own claims from the perspective of other participants. “The moral point of view,” Habermas argues, “calls for the extension and reversibility of interpretive perspectives so that alternative viewpoints and interest structures and differences in individual self-understandings and worldviews are not effaced but are given full play in discourse” (*RDE*: 58). In short, we cannot know in advance of deliberation which features of human motivation, identity, and psychology should be known by moral deliberators and which should be hidden by a veil of ignorance.

But so far Habermas has only shown us why a monological point of view is unsuitable for moral deliberation. He has not shown us how actual discourses can be insulated from the very strategic motives and power inequalities that he has criticized in contemporary political life. What makes an *actual* discourse an *ideal* discourse? To put the problem another

attractiveness of democratic deliberation as a whole, but rather with how it might correct some of the

way, how are participants in actual discourses to know when they can legitimately criticize or reject the determinate motives, preferences, or beliefs of others and when they must accept others' motives, preferences, and beliefs? Here Habermas notes that the moral point of view of actual discourses rests not in the discourses themselves, but in the presuppositions of argumentation or "*idealizing* assumptions that everyone who seriously engages in argumentation must make *as a matter of fact*" (RDE: 50). The idea here is that any citizen who engages in an argument to convince another of the truth or validity of some claim, or who wishes to come to an agreement about some matter, must make certain assumptions about what she is doing and what standing the other person has in the discourse. If the speaker is sincerely trying to *convince* the other person of the validity of some claim, rather than simply trying to coerce or to deceive that person into acting in a certain way or believing a certain claim, then the speaker must regard that other person as, among other things, capable of taking part in a discourse, as worthy of equal recognition, as free to contest and reject certain claims, and as free to express opinions and beliefs of her own.⁷⁰ As Habermas puts it, "in the symmetry conditions and reciprocity expectations of everyday speech oriented to reaching understanding, there already exist *in nuce* the basic notions of equal treatment and general welfare on which all morality turns" (RDE: 50).

The moral point of view from which we make distinctions between interests and motives which should be accommodated and those which should not, then, rests in the logic of argumentation itself which all participants in practical discourses oriented to understanding always already assume. "It follows from the aforementioned rules of discourse," Habermas reasons, "that a contested norm cannot meet with the consent of the

alleged deficiencies of monological approaches.

participants in a practical discourse unless (U) holds, that is, Unless all affected can *freely* accept the consequences and the side effects that the *general* observance of a controversial norm can be expected to have for the satisfaction of the interests of *each individual*" (RDE: 93). Distinctions between objectionable and unobjectionable interests and motives and decisions about their role in the design and justification of principles of justice thus become a matter of public reasoning subject to the critical point of view of the "ideal(izing) presuppositions of discourse"⁷¹ In the end, then, it is the presupposed universalization principle in practical discourses which "works like a rule that eliminates as nongeneralizable content all those concrete value orientations with which particular biographies or forms of life are permeated" (MCCA: 120). How that universalization principle is to be understood in particular cases is up to the participants, but the general principle which guides those actual distinctions does not depend for its validity on the always less-than-ideal deliberative situation in existing societies. The moral point of view "implicit in the...public practice of argumentation" ensures that participants abstract from their particular preferences and interests so that "the core of generalizable interests can then emerge step by step" (RPUR: 58, 68). "Discourse ethics," Habermas writes, "views the moral point of view as embodied in an intersubjective praxis of argumentation which enjoins those involved to an idealizing *enlargement* of their interpretive perspectives" (RPUR: 57).

⁷⁰ For a more formal and comprehensive set of "rules of discourse" that Habermas regards as the "inescapable presuppositions" any speaker who is oriented to reaching agreement will make see RDE, pp. 86-91.

⁷¹ In an earlier version of this chapter I referred to the critical point of view as the "ideal speech situation", a term which, as Thomas McCarthy has reminded me, Habermas has moved away from in his more recent work. McCarthy suggests that the term "ideal speech situation" now makes Habermas "nervous", presumably because of its naïve sociological implications. But whereas the move to focus on the presuppositions of discourse might make Habermas less sociologically nervous in one sense, in another sense I think that move damages rather than improves his theory when we consider the preconditions necessary for the development of citizens' "postconventional moral consciousness" (see section III below). Nevertheless, I am grateful to Thomas McCarthy for discussion on this point.

So Habermas can respond to concerns about the sorry state of actual discourse in existing societies and its threat to discourses of justification by pointing to the critical role of the idealizing presuppositions of discourse which must be approximated if actual discourses are to be taken seriously as discourses of political justification. Moreover, over time and under the right institutions Habermas believes that citizens will develop a “postconventional stage of moral consciousness” whereby “judgment becomes dissociated from the local conventions and historical coloration of a particular form of life” (*DE*: 109). In that case, citizens would be motivated to act in accordance with principles of justice, not because they fit with any of their private motives, preferences, or interests, but instead because they recognize them as valid. In short, over time the idealizing presuppositions of discourse and the associated principles of universalization (U) and discourse (D) become internalized by deliberating citizens.

Motivational Possibility

Habermas recognizes that while the moral point of view must be assumed by participants in practical discourse, the moral norms that acquire validity through practical discourses may, nevertheless, lack a direct motivating force. That is a consequence, he thinks, of moral discourses having to operate from a more abstract standpoint than the ethical and evaluative discourses which provide individuals and groups with answers to questions about the good life. Ethical and evaluative discourses are concerned with questions whose answers can more immediately motivate citizens’ action because answers to those questions “refer to what, from the first person perspective, is in the long run good for me or for us – even if this is not equally good for all” (*RPUR*: 66). Moral discourses, by contrast, are focused on questions of justice – on what is “in the equal interest of all” in societies where answers to

the ethical and evaluative questions differ – and therefore are a step removed from the wealth of motivational potential vested in ethical life (*RPUR*: 66). Habermas recognizes, then, that “moral judgments, which provide demotivated answers to decontextualized questions, require an *offsetting compensation*” (*DE*: 106, my emphasis).⁷²

The problem here, as with earlier theories, is to show how principles of justice can “stick” (Williams, 1981: 122). That is, to ensure that the citizens who must live according to the endorsed principles are *capable* of adhering to those principles without over-taxing their capacities or over-burdening them with too rigorous demands. At the same time, however, a solution to the problem of motivational possibility should not demand that principles accommodate objectionable motives. For Habermas in particular, the problem is not to show how a certain substantive conception of justice is within the realm of practical possibility because discourse ethics gives us only a procedure to test principles and not a substantive set of valid principles. Instead, Habermas has to show more generally how *any* principle which might be accepted as valid via the discourse ethics procedure is motivationally possible given that justification of principles can rely only on motives and other motivational complements which survive the critical assessment of the practical discourse.

Habermas considers four possible sources of motivational potential:

⁷² Here Habermas seems to regard moral and religious worldviews – comprehensive doctrines, to use Rawls’s term – as imbued with motivational potential. While I agree with Habermas that there is connection between motives and moral and religious doctrines, we should be careful not to equate the two ideas in the discussion about the demands of realism more generally. Like the distinction we noted in the previous chapter between motives and interests, motives and comprehensive doctrines are different in the sense that motives induce an agent to act while comprehensive doctrines, like interests, give an agent a reason to act, but not necessarily the motivation to act.

1. Motivating Force of Reasons

In the first place, Habermas identifies motivational potential in the reasons and arguments circulating in practical discourses. Like Scanlon (1998), Habermas thinks that there is a “weak motivating force” in the very reasons which justify valid moral norms. A person can or should be motivated to act according to the demands of valid moral principles because those principles are supported by good reasons – reasons which, Habermas might remind us, are those which are (or could be) accepted by all participants in a practical discourse.⁷³ A certain “socially integrating force” is derived from the fact that “the addressees of the legal norms may understand themselves...as the rational *authors* of those norms” (BFN: 33). What could be more motivating than the idea that the principles which one is asked to endorse and to support are simply those principles which one has given to oneself in the strong sense of having authored those principles oneself?

But as Habermas admits, the motivating force of reasons alone is “weak.” That weakness, it seems, derives in part from the fact that although citizens give themselves the principles to which they are subject in actual discourses, there remains a gap between the generalizable interests which are served by those principles and their own nongeneralizable interests which concern them and perhaps motivate their own behaviour more immediately. There may be good reasons to endorse a particular norm which reflective, deliberating citizens come to recognize, but those good reasons still have to find a way back to the motivations of often unreflective and non-deliberating political actors going about their

⁷³ Recall the principle of discourse ethics (D) which holds that “only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse*” (MCC.4: 66).

ordinary, daily business. So we need to identify other sources of motivation if the moral norms justified by discourse are to be anything more than utopian hopes.⁷⁴

2. *Strategic and Communication Action*

Another candidate for motivation, then, is the potential fit between strategic and communicative action. Because agents in political discourse are asked whether the consequences of norms are acceptable from their own perspectives, there is an opportunity for agents to consider whether those candidate principles are compatible with their antecedent motives as well as their antecedent interests, preferences, and beliefs. As William Rehg notes, “public discourse is not merely a cognitive exercise but mobilizes reasons and arguments that draw on citizens’ interests, values, and identities. Political discourse thus brings in the citizens’ actual sources of motivation and volition” (1996: xxviii). On this reading, discourse ethics appears to re-introduce elements of Hobbes’ and Gauthier’s approaches insofar as it does not demand that deliberation about candidate principles be insulated from the participants’ antecedent preferences, interests, and beliefs.⁷⁵

Again, however, this source of motivational potential is less promising than it appears when we recall that Habermas requires that determinate interests and preferences be open to critical assessment in practical discourses. Political actors might find a fit between their *generalizable* interests and valid principles, but gaps between *non-generalizable* interests or

⁷⁴ In this Habermas differs from Scanlon (1998), for example, who thinks that the motivational force of reasons is all that a conception of justice need demonstrate. That is, Habermas thinks that the threshold of motivational possibility for a realistic political theory is higher or, to put it another way, that the limits of the practically possible are narrower than Scanlon takes them to be. Scanlon thinks that reasons – or, more precisely, the right reasons – are sufficient, that no additional motivational force, such as a desire to be just or moral, is required. By contrast, Williams (1981, 1985) thinks that we’re already lost if we’re not dealing with people who don’t have a desire to be moral or just at some level. That same intuition seems to motivate Hume (1978, 1983).

⁷⁵ On the relationship between discourse ethics and rational choice see Heath (2001).

preferences and those same principles will likely remain. “However gentle it may be,” Habermas writes, “the force of normative claims will be experienced by actors as externally imposed coercion, unless they make it their own as moral force, that is, unless they convert this force into their own motivations” (BFN: 67). The question remains then: How can the motivational difficulty raised by the apparent gap between justified principles and citizens’ motivations be repaired without violating the integrity of the principles which are deemed valid in discourses of justification?

3. The Motivating Potential of the Lifeworld

While some motivational capital is available in the fit between generalizable interests and principles, on the one hand, and in the “weak motivating force” of reasons on the other hand, Habermas knows that a richer and more immediate source of motivation is necessary for principles validated by discourse ethics to be more than principles of a naïve utopianism. The third potential source of motivational possibility, then, is in the relationship between the lifeworld and the discursive moral point of view. If a fit can be found between the values and practices of citizens’ ethical conceptions of the good and the moral point of view, then a more reliable source of motivation might be found.⁷⁶ But even here Habermas recognizes problems that have to be overcome. “Within the horizon of the lifeworld,” he writes, “practical judgments derive both their concreteness and their power to motivate action from their inner connection to unquestioningly accepted ideas of the good life, in short from their connection to the ethical life and its institutions” (DE: 108-9). However, he continues, “the abstractive achievements required by the moral point of view” pose a “risk [to] all the assets of the existing ethical substance” (DE: 109). So if motivational capital is to be found in the

relationship between the lifeworld and the moral point of view it certainly cannot be a direct relationship given the differentiation between moral and ethical questions that the moral point of view brings with it.

4. *Moral Consciousness*

Habermas's answer to the still lingering gap is to introduce a theory of moral psychology which shows how individuals can reach a "postconventional stage of moral consciousness" whereby they are motivated to act on the basis of rational principles rather than by self-interest or a substantive ethics. "[A] principled morality," Habermas explains, "depends on socialization processes that meet it halfway by engendering the corresponding agencies of conscience, namely, the correlative superego formulations....Aside from the weak motivating force of good reasons, such a morality becomes effective for action only through the *internalization of moral principles in the personality system*" (BFN: 113-4). This seems to take us back to the first source of motivational potential – the motivating force of reasons – but unlike the initial description of the motivating potential of reasons, Habermas tries here to show just how the "weak motivating force" of reasons can be embedded in a moral psychology.

Relying on the theory of moral development defended by Lawrence Kohlberg, Habermas maintains that discourse ethics is not beyond the limits of the practically possible because the stage of psychological development it requires is within the range of human capacities. And that stage of psychological development, Habermas thinks, can emerge where all citizens are engaged in practical discourses and placed in situations which encourage them to recognize the contingency of their own preferences and interests, and

⁷⁶ This assumes that citizens are motivated by the ethical conceptions of the good. I will address this

especially their own ethical beliefs. As Heath describes it, “[w]hereas sociologists have traditionally sought to introduce some shared motivational structure – a norm, value, role, etc. – into the analysis of interaction that would directly influence what agents decide to do, Habermas begins by abstracting from immediate problems of decision, in order to determine what interactive structures must be in place for agents to develop the cognitive and linguistic resources needed to deliberate rationally about action” (2001: 26). By engaging in discourses and recognizing the role of the moral point of view in those discourses, citizens makes those crucial first steps towards having the content of their ethical beliefs and culture meet discourse ethics “halfway.”

Yet, in the end, this strategy also fails for Habermas, although the reasons are more complicated and, for that reason, take up the remainder of the chapter. I argue that Habermas’s attempt to locate motivational potential in a theory of psychological development is undermined by his inattention to the substantive preconditions of ideal deliberation, on the one hand, and by the shift from *real proceduralism* to institutional or *structural proceduralism* made in *Between Facts and Norms*, on the other. In the next section, then, I argue that the development of the postconventional stage of moral consciousness which Habermas regards as necessary to demonstrate the motivational possibility of discourse ethics will remain a naïve utopian wish unless his procedural conception of democracy is supplemented by a more substantive account of justice. That is, unless Habermas is prepared to acknowledge that his conception of democracy must recognize and develop its substantive assumptions, the moral consciousness which is supposed to emerge from and support that conception will fail to do so. In section III, I argue that Habermas’s move in his later work towards more impersonal and macro-level models of discourse, along with an

assumption in Chapter 5, “Reasonable Pluralism.”

attempt to “unburden” citizens of the motivational problem by relying on a legal system with sanctions for non-compliance, threatens the mechanism by which transformations of citizens’ pre-political preferences and interests can be achieved. Moreover, that move threatens to undermine the mechanism which helps citizens to develop and act from the “postconventional stage of moral consciousness” – that is, the psychological disposition to be motivated to act on principles because they are valid and not because they fit with some pre-political preference or interest that the agent may have.

II. Procedure and Substance

Although there are some substantive norms of justice implicit in Habermas’s theory, Habermas seems not to regard those substantive norms as clear principles of justice like those offered by Rawls in his justice as fairness. But without clearly articulated substantive principles of justice given prior to democratic deliberation, it is hard to see how Habermas can insulate deliberation from objectionable preferences and inequalities and how the necessary postconventional stage of moral consciousness can develop. The discursive point of view from which we distinguish between generalizable and nongeneralizable interests, on Habermas’s account, is rooted in an idealized speech situation in which citizens regard each other as equal participants in discourses about principles where the interests of all should count equally. But that formal description of the idealized speech situation must be supplemented with substantive conditions and resources to ensure that participants actually do stand in relations of equality with each other and that objectionable interests and power inequalities are excluded.

Habermas’s critique of the Rawlsian original position relies on the idea that individuals are unable to create the necessary distance between their reasoning about just principles and

their own private preferences. Individual reflection, he argues, is no substitute for collective discourse on permissible motives, preferences, interests and acceptable principles of justice because only in communicative relations with others can we ensure that all relevant information is presented and that private preferences are subjected to collective assessment. In light of that argument, it is difficult to see how deliberators in Habermas's procedural conception of democracy can successfully adopt the attitude of mutual recognition and equality implicit in the discursive point of view and how all individuals can effectively and equally voice their interests and arguments without explicit guidelines on just how that can and should be achieved. Participants in such discourses must have a set of guaranteed rights and freedoms on which they can rely and adequate material resources to ensure that those rights and freedoms can be effectively and equally exercised. Moreover, in the absence of these substantive guidelines and material resources, the postconventional stage of moral consciousness which Habermas holds is necessary to demonstrate the motivational possibility of discourse ethics is not likely to emerge.

To see the problem more clearly, consider again the dispute between Habermas's discourse ethics and Rawls's justice as fairness on the issue of procedure and substance as they relate to the motivational possibility of a conception of justice. Rawls holds that justice as fairness is within the limits of practical possibility because it makes not unrealistic demands of the cognitive capacities of citizens. The sense of justice necessary to motivate citizens to support the conception would develop over time, Rawls holds, once just institutions were in place.⁷⁷ But that means that we must be able to identify substantive principles of justice according to which a well-ordered society would be organized. With his two principles, Rawls not only identifies the set of rights and freedoms necessary for a well-

ordered society, he also identifies the requirements of a fair distribution of resources which would ensure that citizens can make effective use of those rights and freedoms. Thus, if we can imagine the well-ordered society in place, we can imagine how it is practically possible for the sense of justice to emerge in that society given knowledge of the well-ordered society and some not unreasonable assumptions about the capacities of citizens.

At the outset, Habermas seems to follow a similar line of thought. With the right institutional set-up – given by a discursive conception of democracy – we can see how acquiring the postconventional stage of moral consciousness is practically possible for deliberating citizens. That is, in situations where deliberators regard each other as equals and in which only the force of the better argument prevails, citizens can develop that attitude which holds that it is good to follow principles simply in virtue of the fact that they are valid and arise from legitimate procedures. But unlike Rawls who recognizes that the shape and proper functioning of just institutions and procedures must be guided by substantive principles, including a principle which ensures adequate resources for participants to have equal standing in deliberation, Habermas appears to balk at the idea that substantive principles and resources must be identified in advance of discourse, even as his conception implicitly endorses some less precise substantive norms.

Seyla Benhabib notes that Habermas does not regard such balking as objectionable because the ideal speech situation, in his view, is guaranteed in the pragmatic presuppositions of argumentation. That is, even in cases where speakers have unequal resources or where they are motivated by strategic aims, the idealized speech situation is unavoidable if they are to regard the outcome of their deliberations as valid (Benhabib, 1986). The problem here is that if others lack adequate deliberative resources, they might agree to norms to which they

⁷⁷ On the relationship between institutions and the sense of justice in Rawls's justice as fairness see

otherwise would not have agreed if they had, for example, a decent education which would improve their communicative competency. It is simply not enough to say that discourses can approximate an idealized speech situation when that ideal presupposes an equality among speakers that does not develop on its own.

Yet, Habermas defends his rejection of more explicit substantive principles of justice like those offered by Rawls by arguing that to do otherwise would be to violate the democratic ideal itself. Indeed, Habermas regards Rawls's defense of substantive principles as violating the "dogmatic core" of a democratic conception of justice whereby "human beings act as free subjects only insofar as they obey just those laws they give themselves" (BFN: 445). Habermas seems to think that Rawls's substantive principles are worked out in theory by the philosopher alone and that citizens themselves cannot thereby regard them as principles that they have given to themselves. As he puts it,

From the perspective of the theory of justice...[the citizens] cannot reignite the radical democratic embers of the original position in the civic life of their society, for from their perspective all of the *essential* discourses of legitimation have already taken place within the theory; and they find the results of the theory already sedimented in the constitution. Because the citizens cannot conceive of the constitution as a *project*, the public use of reason does not actually have the significance of a present exercise of political autonomy but merely promotes the nonviolent *preservation of political stability* (RPUR: 69-70)

In an effort to avoid that same fate, Habermas opts for a more procedural than substantive conception of justice. If a truly impartial perspective must be located in actual discourses, then substantive principles that are presented prior to such discourses can only be regarded as "contribution[s] of a participant in argumentation" (DE: 66).⁷⁸

Cohen (2002).

⁷⁸ Here, as elsewhere, my claim is not that Habermas's theory in fact eschews substantive principles altogether, only that Habermas *says* that we should eschew such principles in theorizing. As I hope to make clear in the following discussion, Habermas's theory does in fact entail some substantive commitments along the lines of principles of justice, but that Habermas himself is hesitant to

But Habermas's attempt to avoid substantive justice falters in two ways. In the first place, he is wrong to think that Rawls's defense of substantive principles is entirely undemocratic or that, more generally, any conception which defends substantive principles must necessarily be undemocratic. In Rawls's political liberalism, citizens are given an opportunity to consider whether the principles of justice as formulated are principles which they could endorse. After the first stage of justification where principles are selected in an original position, a second stage of justification proceeds whereby citizens are asked whether they could endorse those principles from the point of view of their respective comprehensive doctrines. For those principles to be the outcome of an exercise of political autonomy it need not be the case that citizens give themselves the principles in Habermas's strong sense of citizens being the "authors" of those principles. It is consistent with political autonomy that citizens can identify good reasons for endorsing principles of which they are not the literal and direct authors. What would it take away from the justification of a law which condemns murder, for example, to say that each citizen is not the "author" of that law in the strong Habermasian sense? Substantive principles presented to citizens need not be seen as violating the democratic ideal, then, so long as citizens are given the opportunity, and are able, to identify good reasons for regarding those substantive principles as principles which they could endorse.⁷⁹

In the second place, it is not at all clear that Habermas has avoided his own prohibition against offering substantive principles. He defends his theory as a procedural conception of democracy which leaves the articulation and justification of substantive principles to deliberating citizens. But the conditions he sets out for ideal discourse rely on

endorse the principle that such substantive commitments should be made or at least that they should not be made as he thinks Rawls makes them.

substantive norms of freedom and equality. In order for discourses to approximate the ideal conditions necessary for their outcomes to be regarded as just, those discourses must meet certain substantive demands such as equality between speakers and the freedom of speakers to object and to consent according to the moral point of view. As Cohen (1996) argues, procedural conceptions of democracy unavoidably rest on certain principles of substantive justice because the procedures are designed in ways which would ensure fair outcomes. If there are criteria for fair outcomes prior to procedural design and prior to the operation of those procedures, however, then substantive principles of justice must be implicit in the procedures themselves (Cohen, 1996).⁸⁰ Although in some places Habermas denies that substantive principles are implicit in discourse ethics, in other places he seems to acknowledge the substance: “in the symmetry conditions and reciprocity expectations of everyday speech oriented to reaching understanding, there already exist *in nuce* the basic notions of equal treatment and general welfare on which all morality turns” (RDE: 50). And again, it shows up in the “dogmatic core” of his project as a whole which he describes as a “self-organizing community of free and equal citizens” (BFN: 445, 7).

So while Habermas claims that substantive principles cannot precede actual discourses if the democratic ideal of citizens giving themselves the principles to which they are subject is to be respected, I have argued that not only is Habermas wrong to think that the democratic ideal is violated by a priori substantive principles but that his own theory relies on certain implicit substantive commitments. Substantive principles of justice – such as guaranteed rights and freedoms and an egalitarian principle of distribution – are necessary for *actual* discourses to approximate the conditions of *ideal* discourse which itself is necessary

⁷⁹ For more on this line of thinking see “Reply to Habermas” in Rawls (1996) as well as Cohen (1999, 2003).

⁸⁰ See also Vernon (2001).

for Habermas's postconventional stage of moral consciousness to develop. A *real proceduralist* theory, in other words, cannot be merely *proceduralist* all the way down. An advocate of proceduralism must at some level endorse – even if only provisionally – some substantive principles of ideals if we are to regard the procedures as normatively compelling or defensible.

In the absence of substantively equal conditions in political deliberation, moreover, strategic motives and objectionable preferences will threaten the integrity of principles. And if the postconventional stage of moral consciousness is defined as the psychological disposition to act on the basis of principles, then that moral consciousness, when it develops under principles infected by objectionable conditions, will be a warped moral consciousness. If, for example, a certain principle which preserves some inequality is deemed valid in an *actual* discourse only because some participant has the power to withhold necessary or desired goods from some other participant and thus elicit his agreement through manipulation rather than persuasion, then the development of the moral consciousness to act on such a principle will itself be infected by inequality. That is, participants motivated to act according to principles will in effect be motivated to uphold inequality as a matter of principle. To establish the conditions necessary for the development of a healthy moral consciousness then, Habermas must recognize the substantive principles implicit in his own theory and work up substantive principles of justice which could ensure that actual discourses at least approximate ideal discourses.

Without access to resources for education and participation, and without a fair distribution of such resources, the formal conditions of an ideal discourse will fail to gain substantive efficacy. It is not enough to say that the conditions necessary for an ideal discourse are “always already” implicit in the presuppositions of argumentation because, as

Habermas is well aware, not all participants in a given discourse aim at communicative understanding. Inevitably some will view such discourses strategically and aim to shape outcomes to suit their nongeneralizable interests. We need more than implicit conditions for ideal discourse in the presuppositions of argumentation, we need some assurance that the implicit conditions become explicit – that citizens *know* what the conditions for ideal discourse are – and that they are motivated to support and act according to those conditions.

The question, then, is whether the implicit substantive commitments that Habermas makes include commitments to a fair distribution of resources. Citizens will be better able to participate as equals in deliberation if they have received an education (which should include some kind instruction on how to participate and deliberate fairly and effectively) and only if they have material resources and time enough to take the opportunity to participate (rather than, say, having time only to work to meet their basic needs). Additionally, if citizens have such varying levels of resources that some of them view participation as a means to protecting their resources while others view it merely as a means to get more, then there can be only faint hope for deliberation oriented around the common good or what justice requires. In short, if actual deliberations are to approximate the requirements of an idealized speech situation and if they are to be oriented to deliberation about the common good rather than the satisfaction of private, unshared preferences, then Habermas must deliver a better account of how educational and material resources can and should be distributed. What we need to know, then, is whether Habermas's implicit substantive commitments are substantive enough to overcome concerns about the distortion of discourse.

In some ways, Habermas recognizes the concern. Thus, we find him arguing that,

the requirement of equal legal treatment is contradicted by those inequalities that discriminate against specific persons or groups by actually reducing their *opportunities to utilize* equally distributed individual liberties. Insofar as welfare compensations establish equal opportunities to make equal use of legal powers,

compensation for material inequalities in life circumstances and power positions helps realize legal equality (BFN: 415-6, my emphasis).

So Habermas recognizes that the opportunity for citizens to participate equally in his procedural conception of democracy – and thus to deliberate about appropriate principles of justice – depends on some material preconditions. At the same time, however, he argues that such material preconditions cannot be guaranteed or delivered in advance of the democratic procedure because that would entail “paternalism” or, worse, an unacceptable reduction of justice to concerns of mere distributive justice.⁸¹ Habermas recognizes that some material resources are necessary but holds that exactly which resources are necessary must itself be the outcome of a discourse. But absent a pre-discourse distribution of whatever those resources might be, and assuming that actual discourses occur amongst unequal participants, it is hard to see how the actual discourses will result in fair principles of distribution. Thus, while Habermas recognizes that more substantive principles of justice are necessary, the priority that he gives to democratic discourse raises serious concerns about whether the necessary resources will actually be delivered.⁸²

By contrast, a realistically utopian conception which endorses substantive principles – which recognizes the need for a fair distribution of resources so that discourses can proceed in an appropriately free, equal, and reasonable manner – helps to ensure that citizens can equally and effectively participate in deliberation and it helps ensure that the outcomes of

⁸¹ “The welfare paradigm of law is oriented exclusively toward the problem of the just distribution of socially produced life opportunities. By reducing justice to *distributive* justice, it misses the freedom-guaranteeing meaning of legitimate rights” (BFN: 418).

⁸² On the feminist movement’s demands for a better distribution of both resources and rights, Habermas writes that “*public discussions must first clarify* the aspects under which differences between the experiences and living situations of (specific groups of) women and men become relevant for an equal opportunity to take advantage of individual liberties” (BFN: 425). He recognizes that “institutionally defined gender stereotypes” are “social constructions” and that they should be subjected to discursive evaluation, but he does not seem troubled by the fact that given a social and political status quo in which women cannot participate as equals with men (because they are still

such deliberation are more likely to be fair. And while those substantive principles might lack democratic procedural legitimacy at the outset, they may be able to gain democratic procedural legitimacy once better-situated deliberators reflect on the defensibility of those principles and resource distributions. If deliberators are more equally situated then they might have greater confidence that the outcomes of deliberation are fair and just rather than outcomes which protect the interests of strategically motivated and better situated individuals and groups. And if the participants have reason to be more confident in the fairness of deliberative outcomes, then they have more reason to be motivated to support and to uphold those outcomes for their own sake. Insofar as a better distribution of rights and resources ensures a better approximation of an idealized speech situation, a defensible approach to justice and motivation should try to elaborate an account of fair distribution prior to actual deliberation. That Habermas fails to do so, even as he recognizes the importance of doing so, makes his approach to justice and motivation less attractive.

Thus, a *real proceduralist* approach to justice and the demands of realism falters insofar as it fails to recognize the substantive preconditions of legitimate or normatively defensible procedures, on the one hand, and the need to insulate the deliberations in some way from objectionable features of the social and political status quo, on the other. Rawls's *hypothetical constructivism* suggests one way to escape the real proceduralist's dilemma: By imagining ourselves in an initial situation of equality, it is possible to construct the very substantive principles of justice which provide the preconditions for *real* discourses to approximate *ideal* discourses. Admittedly, the democratic deficits of *hypothetical constructivism* are a problem and the *real proceduralist* view offers not only a compelling account of those deficits, but also some possibilities for remedies. Thus, we begin to see that while neither the *real proceduralist* nor

subject to the, at this stage, un-evaluated stereotypes and barriers) the outcomes of those actual

the *hypothetical constructivist* variants of realistic utopianism are capable of addressing the demands of realism alone, when we begin to combine certain features of each view, a more attractive realistically utopian approach comes into view. But before we flesh out that more attractive realistically utopian view, we should consider another weakness of Habermas's attempt to resolve the problem of motivation, as well as the tension between justice and the demands of realism more generally.

III. Impersonal Discourse and Citizen Motivation

In another move which threatens the plausibility of his solution to the problem of motivational possibility, as well casting doubt on his solution to the tension between justice and the demands of realism more generally, Habermas undermines the conditions necessary for the healthy development of the postconventional stage of moral consciousness by turning away from *actual* discourses to arrangements which could ensure an *anonymous* "flow of communication" and reasons in *Between Facts and Norms*. In that work Habermas claims that a theory of democracy, if it is to be realistic, must confront and overcome the problem of social complexity. The problem, as Rehg notes, is that "the functional complexity of contemporary societies no longer permits direct democratic control but rather requires indirect administrative measures guided by expert knowledge" (1996: xxx). The ideal of a macro-subject or democratic community consciously and reflectively organizing its institutions, economic and social processes, and citizens' behaviour according to the principles of a shared conception of justice is, on Habermas's view, naively utopian. There is simply too much information to take in and to process for any individual citizen, or any group of citizens, to successfully construct and apply a normative model to the institutions

discourses may not deliver a sufficient critique and evaluation of those stereotypes.

and processes of contemporary societies in the direct way dreamt of by “philosophers of justice” such as Kant and Rawls.

Habermas’s solution to social complexity and the “overburdening” of the democratic subject is to move away from the idea of normative steering through more personal democratic discourse to a macro-level, more anonymous, public sphere and to replace the idea of the morally judging citizen with the notion of the legal subject “unburdened” by the cognitive, motivational, and organizational demands of contemporary complex societies (BFN: 114). “Various symptoms of...a *cognitive overburdening* of deliberative politics,” he writes, “lend support to the assumption, by now widely accepted, that discursive opinion- and will-formation governed by democratic procedures lacks the complexity to take in and digest the *operatively necessary* knowledge” (BFN: 320).⁸³ So the discursive-democratic point of view that Habermas reconstructs in his earlier work is transformed in BFN into a “structuralist point of view” including a legal system of sanctions for noncompliance, in order to inject what Habermas thinks is a necessary dose of realism into the empirically naïve philosophy of justice. “It is precisely the discourse-theoretic approach,” he writes, “that introduces a realistic element insofar as it shifts the conditions for a rational political opinion- and will-formation from the level of *individual* or group motivations and decisions to the *social*/level of institutionalized processes of deliberation and decision making” (BFN: 461-2).

Drawing on insights gained from what he calls the “sociology of law”, then, Habermas shifts the normative core of his political theory from actual discourses of all citizens to a

⁸³ Just to be clear on the difference between the issue in this section and that considered in the last section: In the last section the problem was that Habermas failed to show how *actual* discourses could approximate *ideal* discourses without relying on the kinds of substantive principles for which he criticizes Rawls. In this section, the problem moves from a concern for the failure to *approximate*

model of “subjectless communications” in which the institutions of a “proceduralized popular sovereignty” and the anonymous circulation of reasons in the public sphere take priority over inclusive deliberations and reflections of citizens themselves. No longer does the justification of moral norms depend of the direct participation of *all* citizens; rather the justification of moral norms and legitimate law now depends simply on the circulation and filtering of reasons in informal and formal public spheres. Indeed, in this updated discourse theory, “the success of deliberative politics depends not on a collectively acting citizenry but on the institutionalization of the corresponding procedures and conditions of communication, as well as on the interplay of institutionalized deliberative processes with informally developed public opinions” (BFN: 298). What matters now for the discourse conception of democracy is not that *all* citizens engage in actual deliberation about candidate norms or laws, because that would be naïve in the face of social complexity; what matters now is that the informal and formal public spheres encourage a “flow of communication” which could ensure that legitimating reasons – “communicative power” – are “transformed through legislation into administrative power” (BFN: 299).

Recall that in examining the possible sources of motivational potential, Habermas held that the reasons which validate norms in discourses of justification offer a kind of “weak motivating force”. The fact that the reasons are those which participants in discourses have themselves accepted allows them to think that acting on those principles is consistent with their regarding themselves as autonomous. “[W]ithout religious or metaphysical support,” Habermas writes, “the coercive law tailored for the self-interested use of individual rights can preserve its socially integrating force only insofar as the addressees of legal norms may at the same time understand themselves, taken as a whole, as the rational *authors* of those

ideal discourse to a concern for the *disappearance* of inclusive, ideal discourses as the central feature of

norms” (BFN: 33). But even that source of motivation is not enough. Habermas recognizes that being able to regard oneself as the “rational author” of a particular norm carries only a weak motivating force because the abstractness of the norm opens up a gap between one’s private interests and preferences and the generalizable interests which the norm endorses or protects. And Habermas thinks that citizens’ attempts to bridge that gap between the abstract principles and their own understandings and actions will “overtax” their rational capacities (BFN: 34). So, Habermas opts to sidestep the motivational and complexity problems by removing much of the responsibility that citizens might have had for developing a sense of justice – that is, a responsibility to be motivated to obey and to support principles of justice *as* principles – and a responsibility to offer, assess, and accept or reject claims made in discourses of justification. Under the new model citizens have an *opportunity* to contribute ideas and make claims, say as members of social movements or intermediate associations, but there is no corresponding responsibility to make an effort to understand the claims of others. That task falls to institutional processes of “filtering” contributions.

Additionally, Habermas attempts to sidestep or avoid much of the force of the motivational concerns by developing and endorsing a legal system whose task it is to ensure compliance with legitimate law. On Habermas’s view, as we have seen, the philosophy of justice approach of theorists like Rawls lacks an appropriate degree of sociological realism. The construction or articulation of a conception of justice under ideal assumptions, rather than the non-ideal world, fails to account for the fact that some citizens will simply lack the motivation necessary to fulfill the obligations of justice. And insofar as a remedy to that problem is absent, the philosophy of justice approach fails to solve the problem of

Habermas’s conception of democracy.

motivational possibility. Indeed, as some theorists hold, to fail to protect the motivated from the non-motivated is to fail to deliver a defensible conception of justice.⁸⁴ To remedy this deficiency or at least to account for it, Habermas argues that obedience to legitimate law can be ensured through a legal mechanism which imposes sanctions on those who fail to comply with legitimate law. As he writes, “whereas archaic institutions supported by worldviews fix value orientations through rigid communication patterns, modern law allows convictions to be replaced by sanctions in that it leaves the motives for rule compliance open while enforcing observance” (*BFN*: 39). This move to incorporate a legal system with sanctions for noncompliance allows Habermas to “unburden” citizens of some of the motivational responsibility they might have under Rawlsian justice and fairness and its associated need for a sense of justice.

This is not to say that Habermas thinks that a legal system alone can completely eliminate the motivational issue. Indeed, he recognizes that even if a system of sanctions for noncompliance is to function effectively, citizens must regard the legal system itself as legitimate even if they think that particular laws are not legitimate. In that sense, he recognizes that motivational stability cannot be delivered by the system of sanctions for noncompliance alone. Nevertheless, the production of that legitimacy and thus motivational stability in Habermas’s later work, as we have seen, is now the work of an institutional and impersonal circulation of reasons rather than the work of individual citizens themselves. And that, I shall argue, undermines the motivational compliance that is required of actual citizens.

Now whether the institutional and more impersonal model of communication is itself naïve is a problem I will leave to the side. What matters for us is whether the shift from

⁸⁴ See Chapter 1.

agent-level discourses to a more anonymous institutional-level flow of communication strengthens or undermines the account of motivational possibility of Habermas's political theory and, more generally, whether it undermines the theory's ability to distinguish between demands of realism which should and those which should not be accommodated. The solution to the problem of motivational possibility which Habermas defends, remember, relies on the idea that principles which are deemed valid through ideal discourses can gain the support of citizens once those citizens advance to a postconventional stage of moral consciousness. "At that stage," as Habermas notes, "moral judgment becomes dissociated from the local conventions and historical coloration of a particular form of life" (*DE*: 109). Instead, the individual who advances to this stage of moral development "takes the perspective of a moral point of view" as the perspective which should guide action and judgment (*MCCA*: 129). That is, the individual learns to be motivated by the insights of morality rather than by the private preferences, non-generalizable interests, or personal beliefs that she has. In the recent characterization of his theory, Habermas claims that "the socially integrating force of solidarity, which can no longer be drawn solely from sources of communicative action, must develop through widely diversified and more or less autonomous public spheres, as well as through procedures of democratic opinion- and will-formation institutionalized within a constitutional framework" (*BFN*: 299). But can that solidarity actually develop under these new conditions?

The problem with Habermas's move away from actual discourse and to "autonomous public spheres" and "subjectless communication" is that no longer do *all* citizens have meaningful opportunities, or a responsibility, to learn how to act in that postconventional and principled manner. Kohlberg's theory of moral development, on which Habermas relies, is a theory of moral *learning*. Agents advance to the postconventional stage of moral

consciousness, if they do, by engaging in practices and activities which encourage that development. In Habermas's more institutional and anonymous conception of democracy, agents are unburdened of the responsibility to take on the perspectives of others, to reflect on the generalizability of their own interests, and to have those reflections and interpretations of their own interests challenged by others in a direct way. Citizens no longer have to experience assessment of their claims in a direct manner and are deprived of opportunities to respond to the assessments in the new framework because in the flow of communication claims are no longer tied to the particular agents who make them.

While some opportunities to contribute to discussion in the informal public sphere are still available to citizens, the decisions made by officials in the formal public sphere now lack the direct link with contributions that they once had (or ideally could have) in the more agent-level discourses defended in Habermas's earlier work. It might be the case that the new institutional arrangements will filter claims appropriately, but that might come at the cost of a widespread apathy among citizens who, we might say, are no longer invited to attend hearings concerning their own interests. An autonomous flow of reasons fails to respect citizens as valuable participants in deliberation about their shared political fate. And when a political system fails to ensure respect for its citizens and their deliberative contributions, it is hard to see how those citizens can regard the principles and laws which the system produces as worthy of their support. Obedience, in that case, might simply be a matter of responding to the sanctions for noncompliance as a matter of self-interested threat aversion rather than obedience to principles and laws that one can regard as those which one has given to oneself in Rousseau's sense.

Habermas might object to this challenge by arguing that not only is the institutional model of discourse more realistic from the perspective of sociology, but that the demands of

motivational possibility are significantly eased under the new model. To be sure, Habermas's more impersonal and institutional characterization of the discourse-theoretic conception of democracy might be attractive precisely because it shows us how we can sidestep motivational and other practical difficulties while retaining the animating ideal of a "self-organizing community of free and equal citizens" (BFN: 7). Habermas holds that "to the degree that practical reason is implanted in the very forms of communication and institutionalized *procedures*, it need not be embodied exclusively or even predominantly in the *heads of collective or individual actors*" (BFN: 341, my emphasis).

If we allow Habermas the assumption that the institutionalized, procedural flow of communication will identify good reasons for principles and laws – reasons that, at the very least, "*could* meet with the approval of all affected" (MCCA: 66, my emphasis) even if they can't claim the *actual* approval of all citizens – then he might be able to defend the characterization of his theory along the lines we used earlier to defend Rawls against Habermas's charge that justice as fairness is undemocratic. That response, remember, held that so long as the reasons which support principles are reasons which citizens *could* regard as good reasons, then a conception of justice need not ensure that citizens are the *authors* of those principles in the strong sense. In that case, however, Habermas would have to give up his criticism of Rawls's justice as fairness, and other similar conceptions of justice, which work up and defend substantive principles of justice in advance of democratic deliberation. But even so, Rawls's approach retains an advantage over Habermas's theory because Rawls calls for a second stage of justification or legitimation in which citizens are asked whether they could endorse the political conception of justice from the perspective of their reasonable comprehensive doctrines. Because Habermas collapses the stages of justification and legitimation into a single model of autonomous communication, citizens are never given

that opportunity to consider whether the reasons offered for principles are in fact acceptable to them.

Moreover, to the extent that the autonomous model of communication no longer requires the inclusion of all citizens in actual discourses, Habermas loses an account of how objectionable interests, motives, and beliefs can be transformed when the moral point of view deems such transformations necessary. The new model may ensure that objectionable preferences, claims, and beliefs are filtered out and therefore have no impact on the principles which are deemed valid, but it fails to show how the citizens who make those claims or who have those interests can be brought to see *why* those interests are objectionable and why changes in their attitudes might be desirable. Indeed, Habermas no longer has a compelling account of how the “interpretation of needs and wants and the change of prepolitical attitudes and preferences” that he praises in Joshua Cohen’s model of deliberative democracy, can actually occur when the deliberations of actual political agents are replaced by an anonymous circulation of reasons (*BFN*: 306).

But perhaps Habermas might hold that such transformations are unnecessary, even dangerous, and that the development of the postconventional stage of moral consciousness is no longer necessary to solve the problem of motivational possibility. If the institutions which ensure the appropriate filtering of facts and reasons are stable and tend to produce principles and laws which citizens could, on reflection, regard as acceptable, then what more does a political theory have to do? I would answer that a political theory does not have to do anything more if the question didn’t rest on the false premise that such institutions can be stabilized once and for all. Once we see that this assumption is naïve, we will also see that Habermas’s attempt to avoid the problem of motivational possibility falls into the utopian

wishful thinking that he wanted to avoid with his more sociologically grounded, but still normatively attractive, conception of democracy.

Institutional Drift and Democratic Citizenship

In an insightful and persuasive essay, Hanna Pitkin argues that the institutions which we establish to ensure just outcomes in our democratic communities have to be assessed and reformed occasionally because over time those institutions and their outcomes tend to “drift” away from the conception of justice according to which they were initially designed. “Most aspects of social life,” she writes, “are left to evolve through drift and private power” (1981: 344). Even if initially just institutions can be established, changes in social, economic and political conditions over time might produce injustices whose correction requires institutional reform. Pitkin maintains, then, that “only citizenship enables us jointly to take charge of and take responsibility for the social forces that otherwise dominate our lives and limit our options, even though we produce them” (1981: 344). Ultimately, the reform of drifting institutions depends on citizens’ capacity to reform those institutions and the existence of a set of appropriate norms which can guide that reform. As Cohen and Rogers write, “it is implausible that the appropriate changes in institutional arrangements will be made unless the norms themselves function as guides to public deliberation” (1995: 39).

With his move to an institutional model of public discourse and a system of sanctions for noncompliance that allegedly unburden citizens of the responsibility to develop the moral consciousness or attitudes necessary to reflect on the justice of their institutions, Habermas undermines the possibility that his conception could remain faithful to his guiding ideal of a democratic society over time. A realistically utopian political theory must show how it encourages the development of citizens’ capacity and motivation to engage in that

kind of reform otherwise it fails to show how justice is possible. At the very least, it has to show how it avoids damaging that capacity. With the institutional and anonymous model of communication, Habermas may think that he can sidestep the problem of motivational possibility and thereby make his conception of democracy more realistic. But as I hope that I have shown here, Habermas has slipped into a naïve utopianism by assuming that institutions can be protected from change and by undermining the development of citizens' capacities to reform those changing institutions when necessary. Indeed, in the latest characterization of his discursive conception of democracy, Habermas seems to suggest that the participation of most citizens is not really necessary for the arrangements to function appropriately. What is particularly striking about that conclusion is how it runs against the grain of the democratic core which inspires Habermas's political theory as a whole. We are left with a conception of democracy which holds that citizens' motivations and capacities are ultimately irrelevant to realizing an ideal in which "human beings act as free subjects only insofar as they obey just those laws they give themselves" (BFN: 445-6).

IV. Conclusion: A More Democratic Realistic Utopia

While Habermas ultimately fails to provide us with a solution to the problem of motivational possibility, it is worth emphasizing the strengths of his view which might help us to design a more attractive realistically utopian approach to the justice and realism tension more generally. One feature of his *real proceduralist* theory is particularly important – namely, the idea that distinctions between demands of realism which should and those which should not be accommodated in principles and norms are to be worked out in a more democratic and deliberative way. Admittedly, much of the work of making those distinctions is done by the point of view of an ideal speech situation rather than participants in discourses themselves,

but the fact that Habermas brings citizens themselves to the centre of debate about such issues is crucial. An ideal speech situation and the presuppositions of argumentation may give us guidelines, but whether particular motives, preferences, interests, and beliefs are permissible or impermissible in discourses about valid principles is something that cannot always be decided by theory alone.

But more than defending discourse as merely a practical necessity, Habermas has shown how democratic deliberation among citizens helps to realize the ideal of a “self-organizing community of free and equal citizens” (*BFN*: 445, 7). It may be misguided to argue that a proceduralist conception of democracy could ensure the realization of that ideal alone, but he recognizes that a political theory which takes the interests and values of citizens seriously must offer a more democratic approach to the realistically utopian challenge. If principles that coordinate the actions of citizens are to be both reliable and just, they must be able to gain the acceptance of those very citizens to whom they apply. And that acceptance, as we have seen, is strongest not when it is the product of bargaining or the mere coordination of pre-political motives, preferences, interests, and beliefs, but when it is granted because principles are seen as legitimate – as serving the generalizable interests of citizens as a whole.

Part II

5

Justice and Pluralism

[P]olitical philosophy is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition.

- John Rawls (1999b: 11)

At the end of the chapter on Rawls, I wrote that two issues with respect to his approach to the demands of realism were still outstanding. One of those issues was the concern that Rawls's approach is not democratic enough and we began to explore that issue by considering Habermas's political theory in Chapter 4. The other concern had to do with how Rawls's theory fares when confronted by the moral and religious pluralism that characterizes contemporary democratic societies. In this chapter, I shall take up the pluralism issue and, in doing so, also say a little bit more about the democracy issue. By the end of this chapter the path should be well-paved to consider a more robust solution to the tension between justice and the demands of realism which I will offer in the final chapter. To get there, however, we need to work through the intersection of justice and pluralism – the other demand of realism on which the dissertation focuses.

The fact of pluralism in democratic societies confronts a conception of justice with what appears to be a dichotomous choice: One can choose to ignore the pluralism of ethical, religious, and philosophical views and insist on independent standards of justice untainted by the demands of pluralism. Or one can adjust the conception to accommodate the fact of pluralism so that principles of justice can play a role in practical political life and thereby avoid the charge of naïve utopianism. Theorists who choose the former option can be criticized for ignoring a permanent and perhaps attractive feature of democratic societies. Moreover, they fail to solve the problem of doctrinal compatibility – namely, that there is

often a gap between what justice demands and what human beings regard themselves as having reason to do in light of their beliefs. Theorists who choose to adjust principles to accommodate pluralism, on the other hand, might succeed in solving the problem of doctrinal compatibility but risk making concessions to unattractive or objectionable features of certain worldviews that we find in contemporary plural democracies. The challenge is to find some alternative to the dichotomy. How can we accommodate pluralism without giving up on the normative and critical value of a shared conception of justice? What exactly is necessary to reconcile a conception of justice to pluralism and, moreover, to what extent (if any) must a conception accommodate the determinate beliefs of citizens?

Rawls (1996) offers an approach to pluralism which relies on the idea of an overlapping consensus of reasonable comprehensive doctrines. In a well-ordered society, Rawls holds, a “publicly recognized conception of justice establishes a shared point of view from which citizens’ claims on society can be adjudicated” (1996: 35). Rawls’s hope is that an overlapping consensus of different comprehensive doctrines on a shared conception of justice can be achieved and thereby accommodate the reasonable pluralism which characterizes democratic societies. And the overlapping consensus is achieved when those who hold different, albeit reasonable, comprehensive doctrines come to endorse the conception for reasons which they find *within their own comprehensive doctrines*. If achieved, an overlapping consensus would ensure not only that a democratic society has a shared conception of justice to which all can appeal in political discourse, it would also offer a solution to the problem of doctrinal compatibility because the reasons citizens would have for endorsing the conception would be found within their own comprehensive worldviews.

We might ask, however, which features of citizens’ comprehensive doctrines should be regarded as fixed points to be accommodated and which should give way to the demands of

the conception of justice to which they are asked to consent? Indeed, given that Rawls introduces the idea of an overlapping consensus as a way to “make the idea of a well-ordered society more *realistic* and to *adjust* it to the historical and social conditions of democratic societies” we might wonder whether political liberalism goes too far in accommodating the sometimes objectionable content of certain existing worldviews (2001: 32. My emphasis). As Charles Larmore observes, we might wonder whether a concern for civil peace trumps basic justice in Rawls’s political liberalism (2003: 385). On the other hand, because Rawls is only interested in accommodating *reasonable* pluralism, we might wonder whether his overlapping consensus and political conception of justice exclude too much of the pluralism found in democratic societies to be practically useful.

In this chapter, then, I defend Rawls’s position that a conception of justice must reconcile itself to the fact and range of reasonable pluralism without resigning itself to the objectionable content of particular comprehensive doctrines. Reconciling justice to the *range* of reasonable pluralism is an appropriate response to an internal challenge of an otherwise attractive conception of justice which secures the justifiability, legitimacy, and doctrinal compatibility of the conception. *Resigning* justice to the objectionable *content* of comprehensive doctrines, on the other hand, would not only threaten the justifiability of the political conception of justice, it would threaten its stability as well. However, I also argue that in order for citizens to reach an overlapping consensus on a conception of justice they cannot rely on deliberation which occurs entirely within their own doctrines, as Rawls suggests. Instead, citizens must be able to introduce some of the content of their doctrines into public deliberation for two reasons: First, because excluding all content of comprehensive views from public political discourse fails to treat fairly those citizens who may have to introduce such content in order to indicate to other citizens the importance of

their claims and thereby receive fairer consideration of their political claims. And second, because we can not always know in advance of public deliberation what should count as a reasonable and what an unreasonable feature of a comprehensive doctrine.

I begin in Section I with brief explanations of the reasonable pluralism of comprehensive doctrines and Rawls's idea of an overlapping consensus which would accommodate the fact or range of reasonable pluralism.⁸⁵ The subsequent two sections articulate and respond to two objections to Rawls's attempt to accommodate reasonable pluralism. One objection holds that Rawls is too eager to accommodate pluralism and criticizes him for adjusting principles of justice for allegedly prudential reasons (Section II). The other objection holds that Rawls is not attentive enough to the pluralism that characterizes democratic societies and suggests that greater accommodations are necessary both for the practicality and justifiability of a conception of justice (Section III). After responding to both these objections I enlist a few ideas from Habermas to guide some revisions to Rawls's approach to accommodating pluralism (Section IV). Finally, in section V, I suggest some ways in which democratic deliberation could improve Rawls's idea of an overlapping consensus.

I. Reasonable Pluralism and an Overlapping Consensus

Rawls asserts that in a well-ordered society a "publicly recognized conception of justice establishes a shared point of view from which citizens' claims on society can be adjudicated" (1996: 35). But modern democratic societies are characterized by a pluralism of religious, moral, and philosophical worldviews which makes agreement on a shared conception of justice difficult to achieve. Moreover, the fact that democratic societies are characterized

⁸⁵ I will say more about the distinction between the *fact* and the *range* of reasonable pluralism shortly.

“not simply by a pluralism of comprehensive religious, philosophical, and moral doctrines but by a pluralism of incompatible yet reasonable comprehensive doctrines,” means that a shared conception of justice will have to accommodate that reasonable pluralism if it wants to meet requirements of both stability and legitimacy (1996: xvii). Reasonable citizens, Rawls notes, are citizens who recognize that the “burdens of judgment set limits on what can be reasonably justified to others” (1996: 61). Though they affirm worldviews which differ from those of other citizens, “reasonable persons will think it unreasonable to use political power, should they possess it, to repress comprehensive views that are not unreasonable, though different from their own” (1996: 60). Faced with this fact of reasonable pluralism, Rawls asks “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?” (1996: xx).

But we might push the issue even farther for Rawls by distinguishing between the *fact* of reasonable pluralism and the *range* of reasonable pluralism. The *fact* of reasonable pluralism holds that there are, among the set of comprehensive doctrines currently held by citizens, a subset of doctrines which are reasonable. But if Rawls think that reasonable pluralism must be acknowledged and perhaps accommodated by a conception of justice, then he must also think that a conception of justice should be designed in a way which would permit the accommodation of other reasonable comprehensive doctrines that may not find expression at this moment in contemporary liberal democracies. In other words, a conception of justice must not only confront the *fact* of reasonable pluralism – that is, the existing subset of reasonable comprehensive doctrines – it must also confront the *range* of

reasonable pluralism – that is, the range of possibly reasonable comprehensive doctrines that may not exist now, but could possibly emerge in a well-ordered society.⁸⁶

In any case, Rawls hopes that an overlapping consensus on a shared political conception of justice *can* be achieved and says that it *is* achieved when citizens who endorse different reasonable comprehensive doctrines accept the political conception of justice for reasons that they find within their own comprehensive doctrines. However, Rawls’s description of the overlapping consensus as a device “to make the idea of a well ordered society more *realistic* and to *adjust* it to the historical conditions of democratic societies,” invites concerns about whether that attempt to accommodate pluralism is a matter of ideal justice or a matter of mere practicality (2001: 32). Rawls himself thinks that while practical concerns should not be dismissed, they are not enough when the main concern should be with what justice requires.⁸⁷ When our concern is with *fair* terms of social cooperation, we should adopt a critical attitude towards existing beliefs and, as we saw earlier, towards existing motives, interests, and preferences. Though we might find that some features of the status quo are not objectionable or even attractive, that discovery should be the outcome of critical reflection on, not an automatic acceptance of, the ethical commitments of human beings and the facts about the world as we now find them.

On closer inspection, then, what appears to be a concern for mere realism and practicality turns out to be a concern with justice itself. In the first place, Rawls is not calling here for an accommodation of the *content* of all existing comprehensive doctrines but instead for a reconciliation of justice with the *fact* of reasonable pluralism – or, as we might say, the *range* of reasonable pluralism. Adjusting a theory to account for a permanent general feature

⁸⁶ Again, I am grateful to David Estlund for making this distinction and for helping me to see its importance to the project as a whole.

of the social world is different than adjusting a theory to incorporate the determinate content of that feature.⁸⁸ To be sure, a theory of justice might have to accept the fact that conflict is a permanent, perhaps natural, feature of social life but it should not automatically accommodate the determinate beliefs of the parties to the conflict. Rawls wants his conception of justice to be political, but political in the right way – that is, capable of gaining the endorsement of a variety of comprehensive doctrines but without making concessions to the power of certain groups and interests in society.

Overlapping Consensus

To accommodate the range of reasonable pluralism using the idea of an overlapping consensus we proceed in two stages. First, we elaborate a defensible “freestanding” conception of justice “working from the fundamental idea of society as a fair system of social cooperation and its companion ideas” (Rawls, 1996: 40). The construction of the freestanding conception at the first stage is insulated from determinate preferences, interests, and beliefs to ensure that it is a fair conception of justice which favours no particular view of the good and which is not simply a bargain struck by unequal participants. With that freestanding conception in place, citizens are asked at a second stage whether they can endorse that conception as a public basis of political deliberation from the point of view of the comprehensive doctrine to which they are attached.⁸⁹

An overlapping consensus is not reached through the give and take of political bargaining, but instead through the unforced decision of free and equal citizens about whether the terms, as given, are acceptable from their own point of view. As Rawls writes,

⁸⁷ “Justice is the first virtue of social institutions...[I]nstitutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust” (1999: 3).

⁸⁸ More on this in section II.

“we do not look to the comprehensive doctrines that in fact exist and then draw up a political conception that strikes some kind of balance of forces between them” (2001: 188). Instead, we articulate a freestanding political conception of justice which draws on ideas implicit in the public political culture of a democracy, but not on ideas taken from moral or religious views, and then ask citizens whether they could support that political conception after reflecting on the content of their own comprehensive doctrines. A political consensus which starts from given beliefs and ethical commitments might achieve a *modus vivendi* – that is, an agreement between parties who regard the agreement as, at that time, in their self-interest. But a *modus vivendi* will be unstable because the parties to that agreement will be ready to abandon it should the distribution of power, their interests, or their beliefs change (1996: 146f.).

The intuition behind the overlapping consensus is shared by Abdullahi Ahmed An-Na'im (1990) who holds that, in the case of human rights, these rights are much more credible and thereby stand a better chance of implementation if they are perceived to be legitimate within the various cultural traditions of the world” (1991: 3). Thus, he proposes that the various cultural and religious traditions engage in a kind of “internal cultural discourse” which works from “the bottom up” in order to achieve legitimacy for human rights and democracy from within each tradition (1990: 3, 7). That is, to achieve an overlapping consensus, the adherents of various worldviews are to consider whether there are resources within their comprehensive doctrines that would allow them to affirm the principles of the political conception of justice. But in searching for such support, the aim is not to try to find shared principles in the sense of locating points where existing worldviews overlap; rather, the aim is to have the adherents of various comprehensive doctrines ask

⁸⁹ On the two-stage approach to justification, see Rawls (1996): 37, 64-66, 140-44 and (2001): 180-1.

whether their worldviews could affirm an already existing, freestanding political conception of justice. As An-Na'im puts it, the process of religious or cultural reinterpretation seeks to find sources of support (legitimacy and stability) for the freestanding conception without altering the *a priori* conception.⁹⁰ So, whereas comprehensive (as opposed to political) liberals might be able to affirm the political conception of justice because it ensures conditions necessary for the pursuit of individual self-development and flourishing, Muslims – the focus of An-Na'im's work – might be able to affirm the conception because it embodies one expression of the Islamic obligation to an ideal of social justice – an obligation given to Muslims in the Qur'an and the Sunna of the Prophet.⁹¹

Of course, the procedure for reaching an overlapping consensus is further complicated once we endorse the idea that a conception of justice should deal with the *range* and not simply the *fact* of reasonable pluralism. Because the *range* of reasonable pluralism might not find full expression in existing societies, there will often be no identifiable citizens or communities who can engage in the kind of internal ethical or religious dialogue which is necessary at the second stage of Rawls's theory construction. Indeed, insofar as Rawls seeks an overlapping consensus among existing reasonable comprehensive doctrines, the possibility exists that a legitimate conception of justice today might not be a legitimate conception of justice tomorrow if a new reasonable comprehensive doctrine emerges which is not already embraced by the overlapping consensus.

⁹⁰ An-Na'im does say, however, that some slight modifications of the *a priori* conception may be necessary, but he is extremely cautious about allowing that possibility. He seems to allow for a little flexibility only to ensure that those who seek to persuade various traditions that human rights and democracy are essential are seen to be acting in good faith.

⁹¹ For a more precise account of how Islam might affirm a political conception of justice see An-Na'im (1990). Additionally, I have been working on a paper titled, "Islam and an Overlapping Consensus" which speaks to this concern.

But it seems that there are two ways to deal with this complication: In the first place, Rawls's idea of an overlapping consensus can accommodate the *range* and not merely the *fact* of reasonable pluralism insofar as it simply amounts to a procedure to accommodate reasonable pluralism rather than a substantive set of principles. That is, because the idea of an overlapping consensus tells us only *how* we should go about accommodating doctrinal pluralism rather than telling us what the content of the consensus should be, Rawls's approach leaves the door open for accommodations of the *range* and not just the *fact* of reasonable pluralism. In the second place, once we make the case for what I call a more democratic overlapping consensus, we will see that the *range* of reasonable pluralism is or can be accounted for because the political conception of justice is left open to adjustments that are accepted by deliberating democratic citizens who are open to the claims made by others as well as to discussions about the reasonableness or unreasonableness of claims and features of existing comprehensive doctrines and any comprehensive doctrine that might be endorsed by some group of citizens in the future.⁹²

Admittedly, if an overlapping consensus is reached, citizens will have different reasons – perhaps even conflicting reasons – for endorsing the public conception of justice. As Rawls puts it,

Citizens have conflicting religious, philosophical, and moral views and so they affirm the political conception from within different and opposing

⁹² More on this in section IV below. I should note that some readers might object that these features of the overlapping consensus do not sufficiently address the problem of accommodating the *range* of reasonable pluralism. Critics might hold that we need to examine all possible reasonable comprehensive doctrines and determine how these are to be accommodated even if they don't currently have adherents. By contrast, I think it is enough to set out and defend a *procedure* for accommodating the range of reasonable pluralism. We need not work through the details of how to accommodate specific comprehensive doctrines until there are actual citizens who endorse those specific comprehensive doctrines and who are prepared to engage in the sort of internal examination of their own worldviews that is an essential part of reaching an overlapping consensus. While it might seem that this response takes me back to the view that only the *fact* of reasonable pluralism must be accommodated, I think I've shown that openness to future accommodations is sufficient for addressing the *range* of reasonable pluralism.

comprehensive doctrines, and so, in part at least, for different reasons. But this does not prevent the political conception from being a shared point of view from which they can resolve questions concerning the constitutional essentials (2001: 32).

Indeed, the conception will benefit from stronger motivational support because the reasons citizens have for endorsing the conception will be drawn from their own moral and religious beliefs and motives. If an overlapping consensus can be achieved then a democratic society will have a shared public conception of justice to guide deliberation, on the one hand, and it will have succeeded in accommodating the fact of reasonable pluralism, on the other. Moreover, the two stage approach ensures that in trying to accommodate pluralism, considerations of civil peace and stability do not trump basic justice because the accommodation is pursued only after a freestanding conception of justice has been described at the first stage of theory.

This way of reaching a consensus on principles of justice, however, faces a number of objections. One objection holds that Rawls's approach to pluralism, which seeks to accommodate reasonable pluralism, risks adjusting the conception of justice to fit with unattractive facts of political sociology and unequal distributions of power. The other objection, by contrast, holds that Rawls's approach to pluralism is not sensitive enough to the actual pluralism of democratic societies and that, in order for a conception of justice to be both practical and normatively compelling, it must aim at more substantial accommodations than those offered by political liberalism. While I think that neither of the objections is fatal to Rawls's theory, we will see that both objections highlight some limits that an attractive approach to accommodating pluralism must respect.

II. An Overlapping Consensus too Accommodating?

The first objection to Rawls's approach to reasonable pluralism that we should address holds that Rawls is too eager to accommodate pluralism and criticizes him for adjusting principles of justice for merely prudential reasons. Proponents of this view argue that bending or adjusting principles of justice to accommodate citizens' particular beliefs – as well as their determinate interests, preferences, characters, and motivations, more generally – whose own defensibility or justification may be in doubt, is unacceptable. Cohen (2003), for example, argues that the Rawlsian method of constructing principles of justice which acknowledges, and sometimes accommodates, certain demands of realism, including pluralism, is misguided.⁹³ Citizens' comprehensive doctrines, may be relevant when we are concerned with the feasibility or "stability" of a conception of justice, Cohen holds, but they are not relevant when the aim is to articulate that initial ideal conception of justice (2003: 199-120). We might regard it as politically prudent to reconcile a conception of justice to the fact of pluralism, but we should recognize that any modifications made to the conception of justice for prudential reasons undermine the attractiveness of that conception because such modifications will entail concessions to objectionable beliefs. So the idea of an overlapping consensus whose aim is to accommodate the fact or range of reasonable pluralism is unacceptable.

One strength of Cohen's objection is that it allows a theory of justice to avoid troublesome, perhaps insurmountable, difficulties in deciding which features of citizens' comprehensive doctrines should and which should not be accommodated. And that means that Cohen protects a conception of justice from the possibility of accommodating

⁹³ My description of G.A. Cohen's position is taken from Cohen (2003, 2003b) and J. Cohen (1993) in which he "extrapolates on some points made by Jerry Cohen in a conversation about Rawls's difference principle" (1993: 271, 288n.7).

objectionable or unattractive features of certain comprehensive doctrines. But that strategy has a significant cost in terms of lost legitimacy and indeed, justifiability, of a conception of justice. If the defenders of a conception of justice say to citizens with minority comprehensive doctrines that their concerns are not deserving even of consideration, let alone accommodation, then it is hard to see how those citizens could come to endorse that conception of justice even if it seems attractive on other counts. If citizens are to endorse a conception of justice and use that to frame their contributions to public deliberation, then we must demonstrate that the conception could be endorsed by someone from the perspective of their own reasonable comprehensive doctrine. Failure to do so would mean that would regard themselves as not having reason to uphold and to live their political lives according to the principles of the conception of justice.

Of course, one might argue that in cases where there is not an automatic fit between the content of citizens' comprehensive doctrines and the demands of a freestanding conception of justice that we should aim instead to change the minds of those citizens. That is, rather than developing elaborate and perhaps cumbersome ideas of an overlapping consensus, perhaps we should encourage citizens to drop their pre-rational religious worldviews and adopt as their exclusive worldview a secular conception of justice. But that option is not open for anyone who thinks that the pluralism of comprehensive doctrines which emerges in democratic societies is *reasonable* – that it is not something to be lamented, but something which is the not unwelcome outcome of the exercise of practical and theoretical reason under free institutions.⁹⁴

By contrast, one might ask why religious toleration is not enough. In the face of religious pluralism why should we aim to achieve an overlapping consensus rather than

⁹⁴ See section III below for more on why “changing minds” is not a suitable option.

simply endorse the idea that comprehensive doctrines are to be tolerated so long as they don't violate the principles set out by the conception of justice. Like the "changing minds" argument, the religious toleration argument fails to recognize the nature of reasonable pluralism, on the one hand, and the need to solve the problem of doctrinal compatibility, on the other. As I will argue below, reasonable pluralism is not simply a brute fact of the world that must be overcome; rather, it is a feature of the world which requires accommodation as a matter of justice and an overlapping consensus provides the appropriate mechanism for that accommodation. Moreover, the religious toleration proposal fails to solve the problem of doctrinal compatibility – that is, it fails to offer an account of how citizens could regard themselves as having reason to support the conception of justice rather than grudgingly resigned to abide by its demands.

In any case, responding to Cohen's objection helps us to see why pluralism – particularly, reasonable pluralism – must be accommodated not only as a matter of political stability, but as a matter of justice itself. The problem with Cohen's objection is that it fails to distinguish between an accommodation of normatively salient or what we might call, to paraphrase David Estlund, "untainted doctrines" (which might be a matter of justice) on the one hand, and concessions to power and objectionable beliefs (which are matters of mere prudence and feasibility), on the other. Not all beliefs and not all features of the social and political world are the product of unequal power relations or are otherwise objectionable. While the fact of *mere* pluralism might be characterized by unequal power between, and unattractive beliefs and interests of, various individuals and groups, the fact and range of *reasonable* pluralism, properly understood, is not tainted in those same ways. The fact of *mere pluralism* describes a social condition in which there are a variety of incompatible and sometimes conflicting comprehensive moral, religious and philosophical worldviews

endorsed by citizens. To accommodate a conception of justice to that brute fact would be tantamount to an accommodation to power.⁹⁵

However, the idea of *reasonable* pluralism, to which a conception of justice should be reconciled, is characterized by two important features: First, the adherents of a subset of all worldviews recognize that it is permissible, indeed natural, for other citizens to endorse other worldviews whose content differs from their own. “[R]easonable persons,” Rawls writes, “are ready to propose, or to acknowledge when proposed by others, the principles needed to specify what can be seen by all as fair terms of social cooperation” (2001: 6-7). Though there may be different views about the good life which arise from the “burdens of judgment”, citizens who are attached to those different views might nevertheless recognize that they should search for and agree to some fair terms of social cooperation which can be shared with citizens who endorse other, perhaps conflicting, moral, religious, or philosophical worldviews. Second, *reasonable* pluralism is not simply a brute fact faced by democratic societies, but a condition which arises, or could arise, when citizens who live under free institutions exercise their capacities for reason and judgment about what constitutes a good life. It is, as I shall presently explain, an *internal* circumstance of justice to which a conception of justice must be reconciled as a matter of justice, rather than an *external* circumstance of justice to which a conception of justice may or may not resign itself as a matter of feasibility.

Internal and External Circumstances of Justice

An *external* circumstance of justice is some fact about human beings or the social world which confronts political philosophy independently of the operation of a principle of justice

⁹⁵ On this point see J. Cohen (1993: 285).

or a well-ordered society. Such external circumstances include such things as scarcity and the limited generosity of human beings about which Hume (1983) famously wrote.⁹⁶ An *internal* circumstance of justice, by contrast, is some fact about human beings or the social world which arises, or would arise, in an already well-ordered society – in a society in which citizens conduct their affairs within institutions that are designed and operate according to fair terms of social cooperation. If we think that some feature of human beings or the social world will develop within that society, then that feature deserves our attention as something that *may* require accommodation as a matter of justice rather than mere practicality.

As Rawls puts it, reasonable comprehensive doctrines, “are not simply the upshot of self- and class interests, or of peoples’ understandable tendency to view the political world from a limited standpoint” as G.A. Cohen might suspect. “Instead,” he continues, “they are in part the work of free practical reason within the framework of free institutions” (Rawls, 1996: 37). When we try to reconcile justice to reasonable pluralism, Joshua Cohen adds, “we are not accommodating justice to an unfavorable condition of human life, since, as the idea of reasonable pluralism shows, we ought not to count moral pluralism itself among the unfavorable conditions” (1993: 272).⁹⁷ Indeed, a circumstance of justice which arises, or could arise, in a well-ordered society where citizens are regarded as free, equal, and reasonable, is a feature of the world which is a serious candidate for accommodation.

That some comprehensive doctrines actually deserve accommodation emerges from the fact that some of the comprehensive doctrines support politically liberal institutions –

⁹⁶ See “Of Justice” in Hume (1983): pp. 20-6. For discussion of Hume’s view and its role in Rawls’s own theories, see Rawls (1999): 109-112 and (2001): 84-5. Rawls distinguishes between “objective” and “subjective” circumstances of justice in these works but I think that the terms “external” and “internal” help to make clearer the point that the fact of reasonable pluralism is a special circumstance.

those same institutions which facilitated, along with the exercise of practical and theoretical reason, the development of those doctrines themselves. And, moreover, the claims of those citizens who affirm those doctrines deserve some kind of accommodation when those citizens are themselves reasonable in the sense of supporting politically liberal institutions and the full inclusion of all other reasonable citizens whose own comprehensive doctrines might conflict with, or even oppose, their own. Citizens are reasonable when they are prepared to recognize other citizens as politically free, equal, and reasonable and when they eschew attempts to impose the content of their worldview on other citizens in political deliberation and decision-making.

G.A. Cohen's failure to distinguish between objectionable and unobjectionable features of the social and political environment leads him to reject the idea that reasonable pluralism can be accommodated by a conception of justice without making concessions to power or objectionable features of particular comprehensive doctrines. Nevertheless, his utopian objection reminds us that when we try to accommodate reasonable pluralism, we must ensure that principles of justice are insulated from power and other objectionable demands of realism. *Reasonable* pluralism, we should remember, is not the same as the fact of *mere* pluralism that we confront in actual democratic societies. Accommodating the former in the right way is a matter of justice because of its internal origin and because it is consistent with compliance to a politically liberal conception of justice. By contrast, accommodating the latter, if we should accommodate it at all, requires careful judgments about the reasonableness or unreasonableness of the comprehensive doctrines that make up that pluralism. That point must be kept in view when we consider the second objection to

⁹⁷ In Rawls's words, "[i]n framing the conception so that it can, at the second stage, gain the support of reasonable comprehensive doctrines, we are not so much adjusting that conception to brute forces of the world but to the inevitable outcome of free human reason" (1996: 37).

Rawls's use of the overlapping consensus to accommodate reasonable pluralism (in the following section) and when we consider some necessary revisions to Rawls's theory (in Section V).

III. Not Enough Accommodation?

Monique Deveaux (2000) offers both practical and normative reasons to defend her claim that pluralism needs to be accommodated much more substantially than it is by Rawls and other liberal theorists. From a practical point of view, Deveaux writes that, "[t]oo often it is the very lack of attention to concrete political claims, developments, and practical restrictions that lead some political theorists to suggest ineffectual and even implausible ways of arranging our political affairs" (2000: 30). The implication is that Rawls's theory, though sensitive to the need to accommodate pluralism in some fashion, does so at such an abstract level that it has no practical use for actual plural democratic societies. Indeed, Rawls's restriction of the comprehensive doctrines deserving of accommodation to those which are "reasonable" ignores a number of other comprehensive doctrines which might not meet the reasonable requirement. Along the same line, Galston writes that "for those who are left out [of Rawls's political liberalism], it is hard to see how liberalism can be experienced as anything other than an assault. Resistance is therefore to be expected, and it is far from clear on what basis it is to be condemned" (1989: 718).

But arguments of mere practicality are non-starters when our concern is with the fundamental principles of justice which should guide institutional design. Galston may be right that "resistance" is to be expected from those who are "left out" of the overlapping consensus and that this will hurt the stability of Rawls's conception of justice, but he is wrong to say that there is no clear basis on which we can distinguish between justified and

unjustified resistance. Justified resistance is the resistance of *reasonable* citizens who endorse *reasonable* comprehensive doctrines that are excluded from the overlapping consensus, whereas unjustified resistance is that of *unreasonable* citizens who are attached to *unreasonable* doctrines. Though we should take seriously resistance of the former sort – because it suggests that those who were left out were left out unjustly – the resistance of the latter should not have an impact on the content of a conception of justice, though that resistance might play a role when we investigate how to *apply* principles of justice to less than ideal political circumstances. Similarly, Deveaux’s complaint that political liberalism might be both “ineffectual” and “implausible” might give us reason to make adjustments in the *application* of principles of justice, but no reason to adjust those fundamental principles themselves.

The claims of unreasonable citizens who are not interested in arriving at fair terms of cooperation with other citizens because they are, instead, only interested in advancing and imposing their particular preferences, interests, and beliefs, might require our attention as a matter of political prudence, but should not be accommodated in the conception of justice itself. A conception of justice which tries to accommodate the brute fact of mere pluralism for reasons of stability alone will soon find itself looking less like a conception of justice and more like a picture of the status quo because it fails to make judgments about which beliefs and practices are tolerable and which are intolerable in a just society. More surprisingly, it will also find that the stability which it took as its guiding aim fails to emerge because it deems the content of all views as worthy of accommodation and therefore offers no guidance to resolve disputes between incompatible views.⁹⁸

⁹⁸ One of the virtues of Deveaux’s approach to pluralism is that it offers a more direct solution to the problem of motivational possibility. By accommodating a wide variety of minority group claims, Deveaux’s conception of justice could more easily locate the motivational support necessary to

Accommodation as a Matter of Justice

In addition to the practical motivation for greater accommodation, however, Deveaux offers a *normative* case for a greater accommodation of pluralism than that offered by Rawls.

Deveaux argues that Rawlsian political liberalism “purchases political legitimacy and stability through a dual strategy of *assimilation* – of culturally diverse citizens to a neutral liberal political culture – and *exclusion*, by requiring that citizens not invoke their particular moral views and identities when debating political fundamentals” (2000: 13). That dual strategy of assimilation and exclusion means that not only will the justice *claims* of minorities not be accommodated, but that they will not even be heard when we construct *principles* of justice and decide on *constitutional* essentials. A conception of justice which ignores the claims of minorities fails to respect the citizens who make those claims and fails to recognize the importance of religious and/or cultural membership to a sense of identity and a life with meaning. On this point I think Deveaux is correct. A conception of justice should recognize the importance of religious and cultural membership to the individuals who are members of those cultures or adherents of those faiths and it should be constructed in ways that ensure the participation of all citizens.

But Deveaux fails to distinguish between the *consideration* of claims, on the one hand, and the *accommodation* of claims on the other. The consideration of a claim requires that individuals and groups are permitted to present, and even encouraged and assisted in presenting, claims in the public political sphere. Moreover, it requires that those claims should be given a fair hearing and assessment in public deliberation. But the

ensure that the conception is stable. But if the strategy for dealing with pluralism is to try to accommodate all or most claims, then it is hard to see what work a conception of justice, as a conception of *justice* as opposed to a conception of stability, has to play in a democratic society.

accommodation of claims should not be as automatic as Deveaux suggests in some places.⁹⁹ At times Deveaux appears to be concerned with the question of *how* minority claims should be accommodated before the questions of *whether* minority claims should be accommodated and *which* claims deserve accommodation have been addressed. While the consideration of claims is necessary to ensure that the fact of reasonable pluralism is itself accommodated in a conception of justice, an automatic accommodation of claims or the content of comprehensive doctrines in principles of justice, absent deliberation about those claims, would be both impractical and unjust. Impractical because the claims of different individuals and groups often conflict and it would be impossible to accommodate all conflicting claims. And unjust because the exercise of political power or the distribution of collective resources should occur only when reasons can be offered with which other reasonable citizens might agree. Automatic accommodation of claims makes a farce of the idea of democratic legitimacy – an idea essential to a defensible conception of justice.¹⁰⁰

Though Deveaux eventually calls for a form of deliberative liberalism in which claims would be presented and assessed, rather than automatically accommodated, the fact that she wants to grant “standing” (2000: 9) to particular cultural minorities *prior to* deliberation might be thought to give greater weight to the claims of those groups prior to the deliberative assessment of those claims. An appropriate response to pluralism should ensure that citizens have standing in democratic deliberation *as citizens*. To suggest that certain groups

⁹⁹ The idea of automatic accommodation of minority claims is evident in a few places in Deveaux’s work: “I suggest that a thicker account of democracy would require the introduction of certain collective cultural rights and special arrangements for cultural minorities and so could help secure a deeper form of accommodation for national and ethnic minorities” (5). “I aim to offer reasons why a more inclusive and deliberative account of democracy...might help us to meet cultural minorities’ demands for social and political accommodation” (6).

¹⁰⁰ To be fair to Deveaux, there are places in her work where she seems to acknowledge that the issue should be one of consideration rather than automatic accommodation, but as a whole, her work

should also receive standing *as cultural groups* appears to give weight to the claims of those groups prior to the presentation and assessment of those very claims. If *after* hearing and assessing the claims of minority citizens the democratic society decides that some special arrangements or accommodations are necessary, then such arrangements should be made. But to make those decisions prior to deliberation would eliminate the very purpose of deliberation about claims before accommodating those claims. Moreover, granting standing to *cultural groups* prior to deliberation might threaten the rights and legitimate interests of some members of those groups who disagree with those who claim to speak for the group. By now we should be aware that within any identifiable group there are often dissenting voices and we should be careful not to weaken those voices with political arrangements that are blind to internal diversity.¹⁰¹

The Ideal of Reasonableness

Nevertheless, we need to consider Deveaux's critique of Rawls's idea of reasonableness itself. Deveaux argues that "Rawls's *ideal of reasonableness* marginalizes, even effaces, the perspectives of those who view their group identity and attendant comprehensive views as inextricably linked to their political convictions" (2000: 91). Certain individuals and groups, Deveaux thinks, are not able to make the distinction between their comprehensive views (about the good life) and their political views (about what can be justified to others). Indeed,

seems insufficiently attentive to the need to clearly distinguish between consideration and accommodation of individual or group claims.

¹⁰¹ I have in mind here the sorts of concerns raised by Okin (1999) and (2005). Specifically, I have in mind the concern that when groups are given standing *as groups* the women who are part of those groups sometimes find that their interests and rights receive worse protection than under political arrangements where citizens receive standing simply *as citizens*. Of course, we should be careful not to let the concerns of internal minorities undermine the legitimate interests and rights of the other group members. The best way to navigate those dangerous waters, it seems to me, is to ensure equal representation of all citizens *as citizens* rather than *as groups*.

she claims that Rawls's "accounts of reasonableness and public reason effectively would prevent certain cultural minorities from discussing political norms and procedures on terms that are acceptable to them" (2000: 93). So she argues for fewer constraints on deliberation and participation in order to adequately accommodate pluralism within a conception of justice. In her mind, "political liberals *unjustly* delimit the kind and scope of arguments citizens may introduce in public debate and so also restrict the social differences we may have reason to recognize politically" (2000: 36).

Though Deveaux may be right that certain individuals and groups will be unable to separate what they regard as *true* from what can be regarded as *politically acceptable*, two considerations speak against allowing citizens to introduce their comprehensive views into public deliberation. In the first place, we should remember that in political deliberation we are often concerned with the distribution of public resources, and that deliberation about justice itself is deliberation about the principles and rules which should guide the adjudication of claims on and distribution of those resources. Arguments about the rules which should guide the distribution of public resources, then, must be presented in ways that could be justified to others who contribute to the stock of public resources and who might have competing claims on those resources. Though some groups might find it difficult to frame their claims in ways that can be justified to others, when the concern is with the distribution of shared public resources, anything less than public justification is unacceptable.¹⁰² To be sure, it is difficult to see how a citizen with a given religious view

¹⁰² Interestingly, many of Deveaux's examples of groups that she thinks are unable to present their claims without appealing to their comprehensive views are groups that want to be left alone in some sense – that is, groups who are not, strictly speaking, making claims for public resources. See, for example, Deveaux's discussion of the Hindmarsh Affair (2000: 95-97). In that example we face a group that denies that a shared political community exists at all in which case we are no longer confronting a group making a claim for accommodation, strictly speaking. See also Deveaux's discussion of the Inuit in Canada's Northeast (2000: 92-3). In that example we face a group which

could endorse a claim made by another citizen with a different, perhaps competing, religious view, on the basis of reasons which appeal to the content of the latter's religion. Allowing citizens to present the content of their comprehensive views might make it easier for them to express themselves, but that move would reinforce, rather than reduce, the problems of moral and religious pluralism in democratic societies.

In the second place, Deveaux seems to conflate the distinction between deliberation about *fundamental principles of justice* and deliberation about *public policies* where just institutions and principles already in place. When the concern is with the fundamental principles of justice, appeals to comprehensive doctrines might be allowed but those appeals can't be expected to provide the justification for the principles and institutions we ultimately endorse. In plural democracies, appeals which rely on parochial moral and religious views fail to accord adequate respect to those who are attached to other moral and religious views and who would not be moved by such appeals. Rawls's justice as fairness "aims at uncovering a public basis of justification on questions of political justice given the fact of reasonable pluralism" (1996: 100). Because it is a conception of justice which aims to be "shared" and endorsed by all reasonable citizens, the justification for that conception must rely on reasons which are "addressed to others" and which could be accepted by others. Faced with reasonable pluralism, those reasons, if they are to be acceptable to others, cannot rest on appeals to the content of un-shared and controversial comprehensive doctrines.

Of course, this leads Deveaux to charge that Rawls regards pluralism as simply a problem to be overcome rather than a positive feature of democratic societies. "[H]is notion of 'reasonable pluralism'," she writes, "simply confirms the *fact* of diverse conceptions of the

accepts some association with the larger political community but which nevertheless seeks autonomous political arrangements. In both cases, the claims presented are not claims of

good but makes no comment as to its desirability. [Rawls fails] to concede the value of social diversity” (2000: 51). While Deveaux may be correct that Rawls fails to recognize the value of social diversity, it is not clear why a theory of justice which aims to be political, not metaphysical, should recognize that diversity as valuable. Rawls regards reasonable pluralism as welcome evidence of successful practical reasoning of citizens under free institutions as well as a challenge to be addressed by a conception of justice.¹⁰³ Rawls’s idea of an overlapping consensus is not an attempt to simultaneously “assimilate” and “exclude”, but instead a sincere effort to accommodate the reasonable pluralism of democratic societies in a way that satisfies the need for mutual justification. We owe other reasonable democratic citizens reasons for our positions and claims which they could understand and endorse without violating their own reasonable comprehensive views. We do not owe them, and a conception of justice does not owe them, some acknowledgement of the value of their comprehensive doctrine to social and political life more generally.¹⁰⁴ The idea of an overlapping consensus on a freestanding political conception of justice, then, attempts to locate consensus on a shared public basis of deliberation that is inclusive of reasonable worldviews and critical of unreasonable worldviews – that is, views which claim exclusive access to truth and which fail to offer reasons in public deliberation that citizens who don’t share their view could accept.

IV. A More Democratic Overlapping Consensus?

Although Deveaux’s critique misses the mark on some points, the idea that we should adopt a more democratic and deliberative approach to accommodating pluralism should be taken

accommodation but demands to be left alone. Neither Rawls’s political liberalism nor Deveaux’s own deliberative liberalism could be expected to provide solutions to such cases.

¹⁰³ See the response to G.A. Cohen in section II above.

seriously. To be sure, the charge that Rawls's political liberalism is insufficiently democratic is a charge that has been advanced quite forcefully not only by Deveaux, but by Habermas.

Habermas suggests that Rawls's approach to pluralism is insufficiently attentive to the content of citizens' determinate worldviews and to the need for a conception of justice which is the outcome of actual deliberation rather than philosophical speculation. As we saw in the previous chapter, Rawls's overlapping consensus and his political philosophy as a whole, Habermas argues, discount democracy in favour of philosophy. Citizens in Rawls's theory,

cannot reignite the radical democratic embers of the original position in the civic life of their society, for from their perspective all of the essential discourses of legitimation have already taken place within the theory; and they find the results of the theory already sedimented in their constitution. Because the citizens cannot conceive of the constitution as a *project*, the public use of reason does not actually have the significance of a present exercise of political autonomy but merely promotes the nonviolent *preservation of political stability* (1998: 69).

Whereas Rawls thinks that locating sources for an overlapping consensus between particular comprehensive doctrines and the political conception of justice should be a matter of discussion for the adherents of those particular doctrines alone, Habermas wants to open those discourses to the political community as a whole. A more public and democratic procedure for accommodating pluralism has two advantages on Habermas's view: First, when citizens are able to introduce content from their comprehensive doctrines into public discourse, there is a greater possibility that they will regard themselves as having reasons to uphold the conception of justice because they would be able to regard themselves as the authors of those principles. Habermas seems to believe that political autonomy is better served, and therefore greater support to be found, in arrangements which allow citizens to develop principles themselves. Rawls's idea of an overlapping consensus, Habermas seems

¹¹⁴ I am indebted to Sarah Song for discussions about the ideas in this paragraph.

to think, entails that the conception of justice described at the first stage of theory confronts citizens as an external or alien constraint on their deliberations rather than a conception which they have given to themselves.

Second, a more democratic procedure for accommodating pluralism provides a means by which the comprehensive doctrines – through which “needs and interests are interpreted” – can be subjected to “criticism by others” (Habermas, 1990: 67). “[T]he revision of the values used to interpret needs and wants,” Habermas writes, “cannot be a matter for individuals to handle monologically” (1990: 68).¹⁰⁵ The idea here seems to be that unless one confronts and engages with points of view different from one’s own worldview, one will, as Will Kymlicka puts it “tend to accept existing practices as givens and thereby perpetuate the false needs and false consciousness which accompany those historical practices” (1989: 898). Because Rawls appears takes the content of comprehensive doctrines off the agenda of public deliberation, he fails to offer citizens opportunities to have their views questioned and assessed by others, and opportunities to explain and to defend their views to others.

As stated, however, both of these reasons fail to justify a more deliberative and more democratic approach to accommodating pluralism. In the first case, it is not clear that political autonomy, and therefore doctrinal compatibility and reasoned support, are improved simply because individuals are able to introduce the content of their comprehensive doctrines into deliberation about a conception of justice. Habermas misunderstands Rawls’s project on this point. Rawls’s political conception of justice is not offered as the content of a constitutional democracy decided upon by philosophers and not citizens. Rather, Rawls’s conception of justice is offered as a shared public basis to guide

¹⁰⁵ As Thomas McCarthy describes Habermas’s project, “rather than contractual arrangements among ‘unencumbered’ individuals with arbitrarily chosen ends, [Habermas’s discourse ethics]

political discourse about the design of fair institutions and policies. Whether citizens actually adopt that conception and use it to guide their political discourse is ultimately up to them. In that sense, citizens retain the political autonomy that Habermas thinks is both normatively essential and practically beneficial. Political autonomy does not require that everyone be the authors of their own laws or constitution in the literal sense; rather, autonomy requires opportunities to assess the terms of social cooperation and opportunities to decide whether one has good reasons to accept or to reject those terms. Rawls's conception of justice is not as undemocratic as Habermas suggests because it is offered, not imposed, as an ideal to regulate political discourse and reasonable citizens always have the final say about whether they could or could not agree to the terms of social cooperation as offered.¹⁰⁶

The second argument Habermas offers for a more deliberative and more democratic approach to pluralism, which appeals to the benefits of public deliberation for evaluating, explaining, and perhaps revising worldviews, also misses the mark. Habermas endorses public deliberation about the content of worldviews because he thinks that such deliberation will allow citizens to be more reflective about, and critical of, features of their own worldviews which “perpetuate false needs and false consciousness” than they would be on their own (Kymlicka, 1989: 898). But, as Kymlicka notes, while we might agree with Habermas that public deliberation might be more likely than internal individual deliberation to encourage reflection on and the revision of unattractive views and practices, it is not clear that such public deliberation has to take place in the formal political sphere (1989: 898). Turning the formal political sphere into a forum of evaluation poses two risks which we should avoid: First, it places citizens at risk of harsh criticism of their worldviews which

involves processes of reflective argument among previously socialized subjects whose needs and interests are themselves open to discussion and transformation.” (1990: xi).

¹⁰⁶ For more on this dispute see J. Cohen (2003).

might damage their self-respect; and second, it risks re-introducing conflict among moral, religious, and philosophical worldviews which are incommensurable but nevertheless reasonable. In short, transforming the public political sphere into a forum for the evaluation of the worth of various worldviews undermines efforts to accommodate a diversity of views and ways of life in the first place. Such a move would effectively shift us away from an attempt to accommodate pluralism and towards a utopian attempt to eliminate that pluralism through deliberation and evaluation.

While Habermas fails to make the case for a more democratic overlapping consensus, the considerations that both he and Deveaux emphasize help us to see two reasons why political liberalism should be more democratic. In the rest of this chapter, then, I shall explain how those two reasons encourage us to adopt an ideal of deliberative democracy within political liberalism.

V. Deliberation and Reasonable Pluralism

1. The Importance of Comprehensive Doctrines

One reason to allow citizens to introduce content from their comprehensive doctrines into political deliberation has to do with issues of mutual understanding and the importance of moral and religious beliefs to the citizens who hold them. Deveaux notes that sometimes appeals to “communities’ traditional beliefs and distinct ways of life in the course of political deliberation...[is] necessary to convey the *nature* and *importance* of the cultural ‘good’ at stake” (2000: 85. My emphasis). On that point I think Deveaux is correct. Many citizens will often fail to take the claims of other citizens seriously because they do not have a sense of how important features of those citizens’ worldviews may be to them. And the failure to take those concerns seriously demonstrates a lack of respect for other citizens. If a democratic

society is to function according to fair terms of social cooperation then mutual respect among citizens must be realized to a sufficient degree. To exclude all content of comprehensive views from public political discourse, then, would undermine one way in which citizens might be able to demonstrate to others that their concerns are, if not justified, then at least important enough that a democratic society should make some effort towards understanding and perhaps accommodating them.

However, while recognizing the need for public deliberation to be more open to the content of comprehensive views for reasons of mutual understanding and equal consideration of claims, we should remember that such content cannot be introduced to *justify* the claims made to others. In democratic societies characterized by reasonable pluralism, justification to others must rely on considerations and reasons which do not depend on nor appeal to any particular comprehensive doctrine. To do so would be unreasonable – that is, it would fail to treat other citizens with respect and, moreover, it would ultimately fail to secure a shared public basis of deliberation. So the introduction of content is permissible to indicate the importance of claims, but should not be regarded as an invitation to use comprehensive doctrines to justify political arrangements and claims on public resources.

In some ways, Rawls recognizes this concern when he introduces the idea of the “wide view of public political culture” which holds “that reasonable comprehensive doctrines, religious or nonreligious, may be introduced in public political discussion at any time, provided that in due course proper political reasons – and not reasons given solely by comprehensive doctrines – are presented that are sufficient to support whatever the comprehensive doctrines introduced are said to support” (2001: 152). So citizens can make contributions to public political discourse through their comprehensive doctrines. But the

final justification of claims and policies must be done in terms of proper political reasons. While the wide view of public reason seems to respond to the objection to Rawls's theory considered above, we should notice that public discourse in this case occurs where a political conception of justice has already been articulated. Indeed, that conception provides the content of the "proper political reasons" that can satisfy the proviso, as Rawls calls it. But our task was to ask whether the demands of realism – in this case the diversity of comprehensive doctrines – provide some justification for changes in the content of the conception of justice itself, not merely to consider whether the content of comprehensive doctrines can be introduced after the content of justice is already settled. In that sense, Rawls's wide view of public reason fails to respond to the objection as stated.

2. *Evaluations of Worth vs. Judgments about Reasonableness*

More importantly, the content of comprehensive doctrines may have to be admitted into public deliberation in order to help citizens distinguish between reasonable and unreasonable arguments and features of their comprehensive doctrines when they aim at an overlapping consensus on a conception of justice. In order for citizens to reach such a consensus on justice they cannot rely simply on deliberation and discussion which is internal to their own doctrines, as Rawls suggests.

While Habermas suggests that deliberation should focus on *evaluating the worth* of various features of citizens' different worldviews, we might instead view deliberation as a means for citizens to *reflect on the reasonableness* of their worldviews. Rawls wants citizens to consider whether they could support the freestanding conception of justice from within their own worldview but we might wonder how those citizens are to know whether certain features of their worldviews are reasonable or unreasonable and how they might be able to

revise their views in more reasonable directions. In the absence of some kind of democratic and deliberative engagement with citizens who endorse other worldviews, one might not take a reflective and critical attitude towards the reasonableness of one's own worldview. All content might be seen, from one's own perspective, as fixed and reasonable. In public discourse with citizens of other faiths and moral views, however, we might begin to see what features of our own worldviews might be unreasonable and what, for that matter, might be changeable. Indeed, in advance of deliberation with other citizens who do not share one's moral or religious worldview, how could one know whether one's view and the sorts of reasons one might offer in public deliberation are reasonable?

Habermas's move to permit *evaluations of the worth and truth* of comprehensive doctrines in the public sphere goes too far. To be sure, subjecting citizens' moral and religious views to the evaluative gaze of others who neither understand nor are moved by those views would fail to treat those citizens with respect. So long as citizens accept the fact and range of reasonable pluralism and offer reasons for principles of justice, political arrangements, and public policies which other reasonable citizens could endorse, then there is no need for Habermasian evaluations. At the same time, even Rawls's wide view of public reason doesn't go far enough towards meeting the demands of democratic pluralism. Again, the wide view permits appeals to comprehensive content in democratic deliberation only where the content of justice is already settled. Moreover, Rawls's wide view of public reason, though it would allow comprehensive content provided the proviso is satisfied in due course, does not seem to leave room for discussions of the reasonableness of the worldviews themselves. That is, Rawls allows comprehensive content as part of citizens' arguments for particular policies (so long as the proviso can be satisfied), but he does not make room in public deliberation for citizens to discuss the prior question of whether the content of their

comprehensive doctrines is reasonable and thus suitable for reaching an overlapping consensus on the political conception of justice.

Between the Rawlsian restrictions on comprehensive content and the Habermasian invitation to evaluation we should locate space for *judgments about the reasonableness* of certain features of comprehensive doctrines and arguments in public deliberation. While Rawls hopes that citizens will come to endorse a political conception of justice from within their comprehensive doctrines, it is not clear that the deliberation leading up to that consensus can realistically and legitimately be conducted solely within the confines of citizens' own worldviews. So while I endorse and defend Rawls's argument for an overlapping consensus of reasonable comprehensive doctrines, I question whether that consensus can be achieved without some *public* deliberation about the content of citizens' various comprehensive doctrines.

VI. Conclusion: Towards Deliberative Citizenship

Whether public deliberation can respect the distinction between evaluations of worth and judgments about reasonableness depends on whether or not *citizens* are capable of understanding and respecting the distinction. One might think that asking citizens to respect that distinction is to go beyond the limits of practical possibility for the average citizen. Moreover, one might think that a political conception of justice which, for the most part, asks that citizens bracket their comprehensive doctrines when they engage in public deliberation asks too much of citizens. Galston (1989), among others, worries that citizens are unlikely to regard themselves as simultaneously politically reasonable citizens or public persons who are prepared to revise their interests, aims, and beliefs as justice requires, on the

one hand and as non-public persons with “aims and attachments” that make up their identities, on the other.¹⁰⁷

In responding to some of the charges against Rawls’s political liberalism, I have relied on the idea that citizens should be prepared, and that they have the capacity, to reflect on and perhaps even revise their determinate motives, interests, preferences, and beliefs when the demands of mutual justification and reasonable pluralism require. Rawls himself relies on the idea of citizens as “reasonable persons” who are “ready to propose certain principles (as specifying fair terms of cooperation), as well as to comply with those principles even at the expense of their own interests as circumstances require” (2001: 191). Appeals to the idea of a reasonable citizen, however necessary, might raise the question of whether it is realistic to rely on such appeals. Can a conception of justice realistically rely on a conception of the citizen which regards citizens as capable of reflecting on and perhaps revising their motives, interests, and beliefs when justice requires? I think that it can, and I will address that question in the next, and final, chapter on deliberative citizenship.

¹⁰⁷ Similar worries are expressed by Sandel (1982).

Justice, Realism, and Deliberative Citizenship

Liberty lies in the hearts of men and women; when it dies there, no constitution, no law, no court can save it; no constitution, no law, no court can even do much to help it. While it lies there it needs no constitution, no law, no court to save it.

- Learned Hand (1952: 190)

Man's capacity for justice makes democracy possible, but man's inclination to injustice makes democracy necessary.

- Reinhold Niebuhr (1944)

The dissertation began with a few central questions: *Why* should concerns about the demands of realism have a role in theorizing about justice at all? *How* are we to distinguish between motives, motivational limitations, and beliefs which should and those which should not be accommodated in a conception of justice? And, finally, how should a conception of justice solve the problems of motivational possibility and doctrinal compatibility? We have seen how Rousseau, Rawls, and Habermas have attempted to offer answers to those questions which avoid concessions to objectionable features of the world and human beings, on the one hand, and simultaneously avoid the pitfall of utopian irrelevance, on the other. Nevertheless, I have argued that while each of these theorists make important contributions to a solution to the tension between justice and realism, none of them offers a completely satisfactory solution to that problem.¹⁰⁸

¹⁰⁸ While Rousseau shows us how citizens' existing motives and beliefs are not necessarily fixed and that political institutions can be arranged in ways that encourage transformation of objectionable features, his *monological* solution to the problem rests on a moral point of view which sits in tension with the fact and range of reasonable pluralism. Rawls, by contrast, offers a *hypothetical constructivist* theory which accounts for reasonable pluralism, but tries to settle too much of the content of justice prior to the democratic deliberation of citizens. Finally, while Habermas offers a seemingly more democratic *real proceduralist* approach to the problem of justice and realism, his solution to the

In this chapter I aim to answer the three central questions by bringing together and more fully elaborating some of the ideas that I introduced and began to develop in the previous five chapters. To solve the first problem – the problem of existing motives and beliefs – we need an argument about which motives and beliefs should be accommodated by a conception of justice and those which should not and, additionally, an argument about why such accommodations are necessary at all. In the case of motives, I shall argue that there are certain motives which should be regarded not as mere practical constraints on the feasibility of a conception of justice, but instead as normatively salient motives which require accommodation as a matter of justice.¹⁰⁹ To identify which motives are normatively salient, I argue that we should adopt the point of view of a society composed of citizens regarded as free, equal, and reasonable – that is, a democratic society of citizens capable of critical reflection on their own and others’ motives as well as their own and others’ interests, preferences, and beliefs.¹¹⁰ Those existing motives which should be accommodated by a conception of justice, then, are those which reasonable deliberative citizens would say are necessary for the protection of their freedom *as democratic citizens*.

In the case of beliefs, I shall argue that a conception of justice should be prepared to accommodate reasonable pluralism. That is, a conception of justice should accommodate the fact that there are a variety of ethical and religious worldviews which, while

problem of motivational possibility fails to protect the mechanism by which citizens can develop the capacities and dispositions necessary to ensure the motivational stability of a conception of justice.

¹⁰⁹ This again raises the concern of David Estlund (see note 8 above) who thinks that when we are dealing with normatively salient motives, the language of “accommodation” doesn’t seem to capture what we’re after. Accommodating normatively salient motives is not really a matter of accommodation, but what justice requires. I continue to use the term accommodation, then, not because I disagree with Estlund’s point; rather, I use the term because it seems a less cumbersome way to refer to the idea that certain motives must count as legitimate considerations in the construction or articulation of a defensible conception of justice.

¹¹⁰ Keep in mind that when I speak of reflection on beliefs, I am referring to the idea of judgments about reasonableness rather than evaluations of worth that I discussed in Chapter 5.

incommensurable with each other, are nevertheless held by reasonable citizens who are prepared to find political principles of justice which recognize other citizens – even those who endorse different and competing comprehensive doctrines – as politically free, equal, and reasonable.

Moreover, assuming that we do not accommodate all existing motives and beliefs and therefore that some competition between the demands of justice and the unaccommodated features will persist, we need a plausible account of how the motivation necessary to support the conception of justice can be generated as well as a plausible account of how doctrinal compatibility could emerge. I will argue, first, that any compelling conception of justice will have to show how it can generate or draw on *internal* sources of motivation rather than rely on *external* constraints, sanctions, or incentives. Moreover, I will argue that a compelling conception of justice must show how it can provide space for the pursuit of reasonable comprehensive doctrines and how the citizens who endorse those doctrines can also endorse the conception of justice from within their own reasonable worldviews. Demonstrating motivational possibility and doctrinal compatibility in those *internal* ways will require two companion ideas: a (realistic) idea of the deliberative citizen and an account of the dynamic relationship between just democratic institutions and the exercise of deliberative citizenship.

In the course of developing these two parts of a complete solution to the problem of realism I will consider and respond to what I have called the realist and the utopian intuitions.¹¹¹ As this chapter will show, both the utopian and realist perspectives on the

¹¹¹ The utopian intuition is best expressed by Mill who, in defending his utilitarian conception of justice against those who charge that the standard is “too high for humanity” and therefore unable to motivate support, replies that “this is to mistake the very meaning of a standard of morals, and confound the rule of action with the motive of it.” “The business of ethics,” Mill asserts is “to tell us what are our duties, or by what test we may know them,” not to accommodate the standard of justice to meet concerns of practical possibility (Mill, 1991: 149). By contrast, from a realist perspective, a conception of justice which has no hope of being realized in this world by human beings as they are,

tension between justice and realism are myopic. While a conception of justice should avoid concessions to objectionable features of the world and human beings, it must accommodate motives and beliefs which are normatively salient. Moreover, while a conception of justice should have solutions to the problems of motivational possibility and doctrinal compatibility, those solutions should retain the emancipatory and critical hope that any conception of justice worthy of the label must retain.

To address the problem of existing motives and beliefs I present the idea of normatively salient demands of realism (Section I) and I introduce the ideal of deliberative democracy as the appropriate point of view from which decisions should be made about which motives and beliefs should be accommodated and which should not (Section II). However, while these two ideas help to solve part of the tensions between justice and realism, they raise two other issues which will have to be addressed – namely, how can we solve the problems of motivational possibility and doctrinal compatibility given that not all existing motives are accommodated at that first stage; and, second, how can we regard the ideal of deliberative democracy as anything but utopian? To address these two issues, I shall introduce the idea of deliberative citizenship, indicate its role in a solution to the problem of realism (Section III), and address the question of whether or not the idea of deliberative citizenship is objectionably utopian (Section IV). Then, in section V, I address the utopian question about why a conception of justice should worry about motivational possibility or doctrinal compatibility at all. I conclude with some thoughts about the difference between pragmatism, optimism, and hope and their connection to justice.

is no more than an intellectual curiosity unsuited for political life. Hume captures that idea when he writes that “custom is the great guide of human life” (1983: 46) and “men must endeavour to palliate what they cannot cure” (1985: 38).

I. Normatively Salient Demands of Realism

Two camps have emerged in response to the problem of existing motives and beliefs: Those in the realist camp hold that a conception of justice which fails to accommodate the existing motives and/or beliefs of citizens will fail to achieve stability. Theorists like Hobbes (1991) and Gauthier (1986), for instance, try to found conceptions of justice on the interests of rational individuals.¹¹² They allege that justice, and the political arrangements which embody justice, will be stable only if the citizens who are bound by that conception of justice have good self-interested reasons for endorsing that conception.

By contrast, those in the utopian camp argue that attempts to base the duties of justice on the antecedent self-interested motives and narrow beliefs of human beings damage the content of justice because those antecedent motives and beliefs are often objectionable.

G.A. Cohen, for example, holds that existing self-interested motives may be relevant when we are concerned with the feasibility or “stability” of a conception of justice, but they are not relevant when the aim is to articulate that initial ideal conception of justice (2003: 199-120).

Attention to existing motives and beliefs, as well as other features of the world and human beings such as interests, inequality, and power imbalances, might be prudent in the application of principles to existing societies, but prudent modifications to the conception of justice itself would undermine the attractiveness of that conception.

Yet, while each camp highlights an important consideration, neither offers an adequate response to the problem. While the realist recognizes the importance of the problem, his solution risks shaping justice in objectionable ways. Moreover, it is not at all clear that an accommodation of all existing motives, for example, which the realists think would be

necessary to ensure stability, is even conceptually, let alone practically, possible. That is, even if we think that a complete accommodation is desirable for reasons of stability, the accommodation of the motives of some citizens may not be compatible with the accommodation of the motives of other citizens. There is no way, for example, to accommodate my desire to be the most respected citizen in my community with the identical desire of another citizen to be the most respected in that same community. We cannot both be the most respected and so there cannot be a complete accommodation of all existing motives. Similarly, it is not at all clear how a conception of justice could, practically speaking, accommodate the existing beliefs of all citizens. What are we to do when one group of citizens refuses to abide by the demands of justice unless they are taken from and explicitly founded upon the Ten Commandments, for example, while another group maintains that all principles of justice must be secular or have secular foundations to gain their compliance. These two demands, based on existing beliefs, simply cannot be satisfied simultaneously in a conception of justice so there cannot be a complete accommodation of all existing beliefs.

By contrast, while the utopian recognizes the dangers of attending to the demands of realism, the utopian strategy of ignoring those concerns risks ignoring those existing motives and beliefs which are, what I call, normatively salient. That is, the utopian fails to distinguish between motivational demands which are objectionable and motivational demands which are unobjectionable or even attractive. Similarly there are beliefs which are objectionable and beliefs which are unobjectionable or even attractive. Or, to put it another way, there are citizens who are prepared to be politically reasonable even when their comprehensive

¹¹² Their success in this regard is dubious as Rousseau's critique of Hobbes suggests and as Gauthier's importation of moral notions like Locke's proviso into the allegedly rational interest explanation of justice shows.

doctrines conflict with other comprehensive doctrines. There is a diversity of features of the world and human beings to which a conception of justice should pay attention and some features which, on reflection, might require accommodation as a matter of justice and not merely as a matter of stability.

Pluralism: Value, Motivational, and Reasonable

It is sometimes held that all of the values which citizens hold dear (or should hold dear) are compatible with each other, that a coherent package of all the concerns that do or should matter to citizens will fit together into a coherent conception of justice. Or, at the very least, it is held that we shouldn't proceed on the assumption that, even if we haven't found such a coherent statement of what justice is yet, it is not possible to do so. Ronald Dworkin writes that

We should hope for a plausible theory of all the central political values – of democracy, liberty, and civil society as well as of equality – that shows each of these growing out of and reflected in all the others, an account that conceives equality, for example, not only as compatible with liberty but as a value that someone who prized liberty would therefore also prize (Dworkin, 2000: 4).

As we saw in the previous chapter, however, agreement on values is not only impossible but undesirable in a society which seeks to protect freedom of thought and conscience. To be sure, the idea of reasonable pluralism suggests that a shared conception of the best life is not only unattainable but undesirable because it could only be achieved through coercion. Moreover, that coercion would fail to do more than make people publicly profess a belief in a shared conception of the good rather than actually believe in such a shared conception.

But perhaps even more importantly, it may be the case that the fundamental political values which one might want to incorporate into a conception of justice, rather than the ethical conceptions that citizens privately embrace, are in tension. Dworkin recognizes that

his view conflicts with that of Isaiah Berlin who holds that “some values may conflict intrinsically” and that “the very notion that a pattern must in principle be discoverable in which they are all rendered harmonious is founded on a false *a priori* view of what the world is like.” Thus, Berlin holds that “the need to choose, to sacrifice some ultimate values to others, turns out to be a permanent characteristic of the human predicament” (Berlin, 1969: li).

What I want to suggest here is *not* that a conception of justice should aim to incorporate value pluralism or value monism. Instead, the appeal to the debate between Berlin and Dworkin is made to remind us that not all the demands of realism that appear to conflict with the demands of a particular principle of justice, are malign. Citizens’ objections to the demands of a particular egalitarian principle of distributive justice, for example, might not be rooted in greedy or jealous motives; rather, the objections might be animated by a concern for liberty – and not just liberty for oneself, but a concern for the liberty of all citizens. That is, in the same way that Berlin recognizes that those things we most value might conflict rather than fit nicely into a coherent whole, I maintain that those features of human beings and the world which might on the surface conflict with other features or human beings and the world and with features of a conception of justice, might nevertheless be features which should be respected and, as far as possible, incorporated into a defensible conception of justice.

The Normative Salience of Reasonable Pluralism

We came across one example of a normatively salient demand of realism in the previous chapter – namely, the set of reasonable moral and religious doctrines which citizens might endorse while acknowledging the legitimacy of other doctrines which are incompatible with

their own. What makes that fact of reasonable pluralism normatively salient are two things: First, that the pluralism of reasonable doctrines is something which would emerge under already just institutions which protect freedom and citizens' opportunities to engage in practical reasoning. And, second, that those citizens who endorse reasonable moral or religious views are reasonable citizens – i.e., individuals who are prepared to acknowledge the legitimacy of other reasonable views and prepared to endorse principles of justice which would be acceptable to other reasonable citizens. In short, what makes reasonable pluralism normatively salient is its origin in freedom and its compatibility with reasonable citizenship.

A conception of justice which seeks to accommodate reasonable pluralism, then, is not making a concession to an unattractive or objectionable feature of the social and political status quo, but instead reconciling itself to the exercise of freedom and practical reason. This is not to say that any and all consequences of the exercise of freedom would require accommodation in a conception of justice; rather, it is to say that some consequences of the exercise of liberty by citizens will be compatible with the legitimate interests of all other citizens in locating shared principles of justice to guide social and political life. Moreover, once we move towards a view of justice as the outcome of an agreement that reasonable, reflective citizens could make, we begin to see how the accommodation of certain features of the world – features which arise from reasonable exercises of freedom – would have to be made if all citizens are to agree to a given conception of justice.¹¹³

Normatively Salient Motives

Recent discussions about the appropriate level of application of Rawls's difference principle – his second principle of justice – provides another way to make sense of the claim that

some demands of realism require accommodation as a matter of justice and not simply as a matter of political prudence. Recall G.A. Cohen's objection to Rawls's application of the difference principle to the basic structure of society rather than to citizens' personal behaviour or the social ethos. Because Rawls restricts the application of the difference principle to the basic structure of society, Cohen argues, citizens are permitted to make choices that violate the spirit of the difference principle itself. Indeed, Cohen maintains that not only does Rawls *permit* incentive-seeking behaviour, but that Rawls must regard that behaviour as not unjust. If correct, Cohen's claim means that justice as fairness depends on an unattractive concession to an objectionable motive – namely egoism. The talented agents who demand incentives, Cohen alleges, do so not because the incentives are necessary, but simply because those agents want the incentives and have the power to demand them successfully. In that case adjusting the content of justice to fit with citizens' antecedent motives would be nothing more than a violation of justice itself for mere practical considerations.

But as the discussion in Chapter 3 demonstrated, applying the difference principle to the basic structure of society and therefore permitting citizens to make choices about their personal behaviour, need not be construed as a concession to objectionable motives like greed or egoism. Instead, that decision about level of application can be seen as a way to make space for motives which are, on reflection, the sorts of motives that an attractive conception of justice might want to protect. There are not only “untainted motives”, such as affection for one's family, that we might want to incorporate into a conception of justice or, at the very least, acknowledge as deserving of some operational space within or outside of the demands of justice (Estlund, 1998: 102, 108). There are also motives that are not merely

¹¹³ Of course, this prompts G.A. Cohen's question about what it is that the idea of agreement adds to

“untainted” but should be positively valued by a conception of justice. Citizens might be motivated by a concern for their freedom – a desire to lead a life free of unnecessary fetters and a desire to have the freedom to conceive of and to pursue a conception of the good that one chooses for oneself as opposed to one that is imposed.

The idea of normatively salient motives and normatively salient demands of realism, more generally, then, responds to both the realist and utopian strategies for dealing with the problem of realism. In contrast to the realist who holds, for example, that existing motives must be accommodated in order to ensure motivational stability, my view holds that accommodations by a conception of justice should be considered only for those unobjectionable motives that a conception of justice should incorporate *as a matter of justice* – in particular, citizens’ motivation to be free.¹¹⁴ Unattractive motives such as greed, egoism, or an appetite for power might be salient when we are concerned with the application or stability of principles of justice, but not in the design of the principles themselves. In contrast to the utopian who holds that we should pay no attention to motivational concerns when articulating principles of justice, then, I hold that there are certain motives that demand our attention and perhaps accommodation as a matter of justice. Here we are not talking about making accommodations for citizens who through weakness of will or malign intent seek to dodge the demands of justice. Rather we are talking about the possibility that citizens who are otherwise committed to justice and fulfilling its duties find themselves pulled in different directions by reasonable demands and motives.

the justness of principles of justice. I will address that question below.

¹¹⁴ Of course a conception of justice cannot guarantee full freedom for all citizens because each citizen’s freedom must be limited when necessary to protect an equal degree of freedom for other citizens. And, moreover, restrictions on freedom may be necessary to ensure that a conception of justice satisfies other interests and concerns such as equality, reciprocity, and publicity. How those concerns all fit together in a conception of justice is an important question; however, my aim is only to demonstrate that certain motives are among those things which must be reconciled.

But while certain motives and reasonable pluralism may be normatively salient and therefore in need of accommodation as a matter of justice, we must have some way to distinguish between those normatively salient motives and reasonable comprehensive doctrines and other motives and doctrines which are neither normatively salient nor reasonable. To meet that need, I suggest that theories of justice should be guided by two ideas: An ideal of deliberative democracy and a conception of deliberative citizenship. In discussing these two ideas, we will also see how addressing the problem of existing motives inevitably raises the problems of motivational possibility and doctrinal compatibility to which the idea of deliberative citizenship offers the first step towards a solution.

II. An Ideal of Deliberative Democracy

As we have learned from our examination of the theories of Rousseau, Rawls, and Habermas, distinctions between demands of realism which should and those which should not be accommodated in principles of justice should be made from a moral point of view. While each of these theorists offer candidates for an appropriate point of view, we have seen how, for different reasons, none of them offers a point of view completely suited to resolving the justice and realism tension.¹¹⁵ Thus, we need a more suitable perspective from which to make the necessary distinctions. Moreover, as Chapter 5, on reasonable pluralism, demonstrated, even when we can identify a normatively salient demand of realism, exactly how that demand should be accommodated remains an open issue. Thus, in order to distinguish between appropriate and inappropriate demands, and to provide a means by which accommodations can be considered, I suggest that we should adopt the point of view of an ideal of deliberative democracy.

¹¹⁵ See note 110 above.

As Samuel Freeman writes, “the ideal of deliberative democracy says that in voting it is the role, perhaps the duty, of democratic citizens to express their impartial judgments of what conduces to the common good of all citizens, and not their personal preferences based on judgments of how measures affect their individual or group interests” (2000: 375). To assist citizens in achieving that aim, or fulfilling that duty, political arrangements are to ensure that there is a free exchange of ideas and reasons, equal opportunities for citizens to make and to assess claims, and equal rights, resources, and capacities to effectively use those opportunities, among other things.¹¹⁶ Moreover, as Freeman and others note, deliberation is to be focused on the common good of citizens as a whole and not myopically on their own antecedent private preferences and interests: “Deliberation aims for a rationally motivated *consensus*, to find reasons acceptable to all who are committed to acting on the results of free and reasoned deliberation by equals” (Freeman, 2000: 390).

The particular motives and beliefs, as well as preferences, interests, and other things that citizens hold prior to the adoption of a conception of justice are a mixture of malign, benign, and normatively salient motives, beliefs, preferences, and interests. Democratic deliberation is a particularly promising way of distinguishing between the malign and the benign features because of the way it requires citizens to address their claims to other citizens, to submit to public scrutiny of their claims, and to justify their claims by offering reasons which other citizens could accept. As Joshua Cohen writes, “[w]hen properly conducted...democratic politics involves *public deliberation focused on the common good*, requires some form of *manifest equality* among citizens, and *shapes the identity and interests* of citizens in ways that contribute to the formation of a public conception of common good” (1997: 69). Indeed, deliberative democracy is expected to make participants more other-regarding in

¹¹⁶ See Cohen (1997), Gutmann & Thompson (1996, 2004), Habermas (1990, 1996), Scanlon (1998).

their behaviour and in the definition of their interests. Participants in deliberation learn to frame their contributions in ways which could reasonably be accepted by other free, equal, and reasonable participants.¹¹⁷ As Cohen continues,

[a]ssuming fair conditions of discussion and an expectation that the results of deliberation will regulate subsequent action, the participants would tend to be more other-regarding in their outlook. The structure of discussion, aimed at solving problems rather than pressuring the state for solutions, would encourage people to find terms to which others can agree. And that would plausibly drive argument and proposed action in directions that respect and advance more general interests (1996: 113).

If, as Rousseau believes, individual citizens' weaknesses, vices, and prejudices seem to "shrink...the limits of the possible in moral matters" (1978: 3.12.2) then perhaps democratic deliberative arrangements can enlarge the limits of the possible by subjecting those weaknesses, vices, and prejudices to public scrutiny.

If we use the ideal of deliberative democracy as the appropriate point of view from which to identify malign, benign, and normatively salient demands of realism, that means that only those demands which could be justified to other citizens as reasonable demands should be considered as candidates for accommodation in principles of justice. And we might expect that those demands which should be accommodated by a conception of justice – that is, those which would pass the ideal deliberative test – are those which reasonable deliberative citizens would say are necessary for the protection of their freedom *as democratic citizens*. Presumably, citizens will be motivated not simply by determinate, contingent preferences but also by more fundamental interests. A democratic citizen, one might expect, will want a set of rights and resources necessary to form and to pursue a conception of the good as well as the rights and resources necessary to participate as an equal in democratic political life. The ideal of deliberative democracy, then, helps to distinguish between the

¹¹⁷ See, for example, Rawls (1996) and (1999c).

nonessential contingent preferences and beliefs, on the one hand, and the more fundamental interests, motives, and beliefs. Citizens could reasonably justify to others the need for principles of justice which would accommodate the more fundamental interests of democratic citizens, but may not succeed in justifying the need for principles of justice to accommodate other antecedent preferences.¹¹⁸ To be sure, the deliberative democratic point of view was what helped us to see why citizens' reasonable moral and religious views, as well as their interest in freedom, constitute normatively salient demands of realism which require some kind of accommodation.

It is important to keep in mind that adopting an ideal of deliberative democracy as the appropriate point of view does not mean that principles of justice are to be the outcome of *actual* deliberations, at least not directly. Actual deliberation, as we learned from our discussion of Habermas, rarely, if ever, approximates the *ideal* of deliberative democracy and risks the possibility that principles of justice will simply reflect the inequalities and other objectionable features of the social and political status quo. The utopian objection to our attempt to solve the problem of justice and realism would carry the day here if we relied on actual, non-ideal, deliberation to make decisions about the content of justice. Indeed, actual deliberation that fails to meet the requirements of the normative ideal of deliberative democracy threatens to descend into mere self-interested bargaining, the outcomes of which could not be regarded as principles of justice to which all citizens could agree.

While the ideal of deliberative democracy offers a first step towards solving the tension between justice and the demands of realism, then, three other questions and concerns now arise: First, how can we solve the problem of motivational possibility given

¹¹⁸ I have not said that citizens will not succeed in defending more particular preferences to other citizens. A moral point of view which ruled out that possibility at the outset would fail to treat

that not all existing motives are accommodated at the first stage of the solution? Second, how can we regard the ideal of deliberative democracy as anything but utopian in light of citizens' capacities and the comprehensive doctrines which they endorse? Finally, and perhaps most obviously, if we are not talking about *actual* deliberation then how does the *ideal* of deliberative democracy play a practical role in discussions about justice? To address these issues, we need to consider the role, capacities, and limitations of what I call the deliberative citizen.¹¹⁹

III. Deliberative Citizenship

While the ideal or actual practice of deliberative democracy is thought by some theorists to be doing the work of distinguishing between reasonable and unreasonable claims, what I want to argue is that it is not simply the institutions or procedures of deliberative democracy but an associated *exercise of deliberative citizenship* which achieves that aim. In short, the capacities, dispositions, and actions of citizens play an essential role in sorting through concerns about motives, beliefs, and the demands of realism of realism more generally. And, moreover, I want to suggest that a certain kind of citizen – the deliberative citizen – is the kind of citizen that we ought to, and can, rely upon for solutions to the problems of motivational possibility and doctrinal compatibility.

citizens and their interests, beliefs, and motives with respect – it would, in effect, risk the totalitarian scenario about which Talmon (1952) and Popper (1966) complain. See chapter 2.

¹¹⁹ In earlier drafts of the dissertation I used the terms “reflective citizen” and “reasonable citizen” to label the conception of the citizen I have in mind here. I now use the terms “deliberative citizen” and “deliberative citizenship,” which I think better capture the spirit of the conception. I first came across the terms “deliberative citizen” and “deliberative citizenship” in Mendelberg (2001). I also thank Loren King who suggested that term in private conversation.

The Capacities and Role of Deliberative Citizens

A deliberative citizen is a citizen who is prepared and capable of reflecting on her own and others' interests, motives, beliefs, and claims in an impartial fashion, prepared to offer reasons that others could accept to justify their claims, and prepared to revise her preferences, interests, and beliefs when required. This conception of deliberative citizenship requires, then, that citizens be more than the self-interested, instrumental reasoners of Hobbesian political thought – that is, citizens who are capable of more reflective reasoning than what would be necessary for strategizing to satisfy one's antecedent aims and preferences whatever those might be.

The conception of deliberative citizenship draws in part on the idea of “civic consciousness” advocated by Cohen and Rogers. They write that “civic consciousness...require[s] a general recognition of the norms of democratic process and equity [i.e., popular sovereignty, political equality, and distributive fairness], and a willingness to uphold them and to accept them as fixing the basic framework of political argument and social cooperation” (1995: 38). Cohen and Rogers embrace a deliberative conception of democratic legitimacy in which the “justification of the exercise of collective political power is to proceed on the basis of free public reasoning among equals” (Cohen 1996: 99). It is this understanding of democratic legitimacy which recognizes a need for the development of “other-regarding” outlooks and behaviour, a “willingness to treat others with respect as equals,” and a greater concern for the “common good.” A deliberative citizen, then, will not only seek to pursue her own aims and interests, but will be prepared to consider not only how her interests are compatible or compete with the interests of others but, more importantly, she will be prepared to consider the common good of a democratic society composed of other free, equal, and reasonable citizens. To put the point another way,

deliberative citizens will be concerned not simply with their own interests, but also with the legitimacy of those interests, with ways to pursue those interests consistent with the rights and interests of others, and with the common good of all citizens.

In a moment I'll indicate why I think that conception of deliberative citizenship is not unrealistic but, first, consider why that conception of citizenship is *necessary*. I suggested earlier that deliberative institutions and procedures alone are insufficient to distinguish between normatively salient and other demands of realism and, I want to add, insufficient for stabilizing a conception of justice once such a conception has been settled upon.¹²⁰ Deliberative citizenship is necessary, then, because it is citizens – not institutions or procedures – who ultimately preserve and abide by the ideal of democratic deliberation and who, ultimately, provide the real motivational and doctrinal stability for a conception of justice.

While it is the case that we want institutions and procedures in place that embody the norms and principles of justice to which we could agree, the initial establishment of those institutions depends on the actions of citizens who seek to establish them. And if just institutions and procedures are to be established, citizens will need some way to judge whether these institutions will promote or damage the principles of justice. Judgment must originate in the aims, dispositions, and normative commitments of citizens themselves. When adopting new, and hopefully just, procedures and institutions, then, there is no choice but to depend on the reasoning abilities, motivations, and normative and ethical commitments of citizen.

Moreover, even if initially just institutions can be established, changes in social, economic and political conditions over time might produce inequalities whose correction will

¹²⁰ See also chapter 5, section III.

require institutional reform. “It is implausible,” Cohen and Rogers argue, “that the appropriate changes in institutional arrangements will be made unless the norms themselves function as guides to public deliberation” (1995: 39). But those norms certainly don’t function as “guides to public deliberation” all on their own; rather, those norms emerge as standards of institutional assessment and reform when citizens themselves introduce and appeal to those norms in public deliberation. Speaking about liberty, rather than a conception of justice as such, Learned Hand makes the point forcefully when he says that “[l]iberty lies in the hearts of men and women; when it dies there, no constitution, no law, no court can save it; no constitution, no law, no court can even do much to help it. While it lies there it needs no constitution, no law, no court to save it” (1952: 190). Ultimately, the ideal of deliberative democracy will have to be carried by deliberative citizens – that is, embedded in the dispositions, motivations, and normative commitments of citizens.

It might be thought, at this stage, that what I am arguing for is similar to Rawls’s idea of the sense of justice. “A sense of justice,” Rawls writes, “is the capacity to understand, to apply, and to act from the public conception of justice which characterizes the fair terms of cooperation. . . [and] a willingness, if not the desire, to act in relation to others on terms that they also can publicly endorse” (1996: 19). Additionally, Rawls holds that citizens must be able to “acquire a normally sufficient sense of justice so that they comply with [a well ordered society’s] just arrangements” (1996: 141). In some respects the idea of deliberative citizenship shares certain ideas with Rawls’s sense of justice, but it goes further than the sense of justice. The sense of justice which requires that citizens “comply with” just arrangements and to “act from the public conception of justice” implies that the content of the conception of justice is settled prior to deliberation. Thus, the sense of justice is a disposition to obey and to support an already settled conception of justice. Deliberative

citizenship, however, acknowledges that the content of justice may not be settled before deliberation and that citizens will need the capacity and motivation to deliberate about the content of justice and, thus, to make and assess claims while regarding others citizens as free, equal, and reasonable. Nevertheless, the two ideas share the intuition that the stability of a conception of justice does not rest ultimately in the institutions which embody that conception, but instead in the attitudes, actions, and disposition of citizens.

Deliberative Citizenship, Motivational Possibility, and Reasonable Pluralism

So how, exactly, does the idea of deliberative citizenship help us to solve the problem of practical possibility? The problem of practical possibility – a more general term for the specific problems of motivational possibility and doctrinal compatibility – arises when a gap appears to remain between citizens' existing motives or beliefs which are not accommodated and the demands of a candidate conception of justice. In the face of that gap, an attractive conception of justice needs to show that citizens can come to endorse the principles even when their antecedent motives and beliefs appear to conflict with those principles. And, moreover, it needs to show that motivational possibility and doctrinal compatibility could be generated internally by a conception of justice rather than enforced through external constraints such as sanctions for noncompliance.¹²¹

Deliberative citizens, I maintain, are citizens who have the capacity and the motivation to deliberate about the requirements of justice, to support a conception of justice, and to recognize when social and political institutions need to be reformed. Moreover, they are citizens who are receptive to reasonable transformations of their pre-deliberative preferences, interests, motives, and beliefs when those transformations are required by

justice and do not otherwise conflict with their most fundamental interests as democratic citizens. Part of the reason why citizens would be open to that kind of transformative possibility, I want to suggest, arises from the fact that an attractive conception of justice will strive to accommodate reasonable pluralism. As we saw with the chapter on reasonable pluralism, citizens would be more likely to support a conception of justice and perhaps more amenable to revisions of the determinate motives and preferences if that conception of justice was adjusted to permit citizens to hold reasonable moral and religious worldviews. That is, if the conception of justice were adjusted to accommodate citizens' fundamental interest in pursuing a reasonable conception of the good, then that accommodation would give those citizens an additional reason to support that conception of justice.

Of course, it is not the case that citizens will immediately endorse the requirements of a conception of justice just because it permits the pursuit of reasonable conceptions of the good or provides room for reasonable moral and religious worldviews. Rather, the point is that if the conception of justice can be endorsed by citizens for reasons that they find within their own worldviews, then, over time, citizens would likely be more receptive to the transformative demands of that conception of justice. Most, if not all, reasonable moral and religious doctrines require that citizens forgo the pursuit of some of their more immediate preferences in order to satisfy the requirements of their morality or religion. Because citizens regard those moral and religious demands as their own, a conception of justice which finds support within the worldviews of reasonable citizens can draw on that same reflective and transformative possibility. When the conception of justice is regarded as part of the worldview of a citizen – when an overlapping consensus has been achieved – she will feel the demands of justice not as external constraints or demands, but as internal demands,

¹²¹ See chapter 1, section II and chapter 4, section III on why the sanctions-for-noncompliance view

demands that must be met by one who wishes to be a faithful adherent of her particular moral or religious doctrine.

Thus, deliberative citizenship, which is necessary for the stability of a conception of justice, *could* be endorsed as a correct model of citizenship by citizens from within their own comprehensive doctrines. That is, citizens can find reasons to be deliberative citizens, and thereby help to support a conception of justice which relies on a conception of deliberative citizenship, from within their own moral and religious worldviews. That we might reasonably *expect* deliberative citizenship to be endorsed from within various comprehensive doctrines depends on whether citizens can be encouraged by institutional arrangements to engage in discourses about their comprehensive doctrines with those who share their worldview.¹²²

In any case, the possibility of deliberative citizenship being endorsed from within comprehensive worldviews offers as reliable a solution to the problem of practical possibility that we can hope for. That solution, of course, does not *guarantee* motivational or doctrinal support in part because it still relies on the not infallible political judgment of ordinary citizens. But it indicates that support is possible and, moreover, it does so without committing the realist mistake of damaging the integrity of principles of justice simply to achieve guaranteed stability.¹²³

fails to deliver stability.

¹²² See An-Na'im (1990, 1992) on the "internal cultural discourse[s]" (1990: 3) necessary to achieve this aim and also for an account of how human rights and democracy might locate legitimacy from within Islamic doctrine. See also Rawls (1999c: 154n.52) on how this might be achieved in Catholicism but also for an account of why a political liberalism should generally avoid developing such accounts from the outside, rather than inside, of comprehensive doctrines.

¹²³ The "could be endorsed" condition is admittedly a weak criterion. But, as I have claimed throughout the dissertation and as I will again argue in the final section, a solution to the tension between justice and realism should be guided by the idea of *possibility* – by an idea of what it is reasonable to hope for – rather than what is *probable* or *likely* because that is necessary for a solution to avoid both the realist and utopian pitfalls that I have argued we should avoid.

IV. Is deliberative citizenship objectionably utopian?

It might still be objected that this solution to the tension between justice and realism is objectionably utopian. It might be said that the ideal of deliberative democracy and its associated conception of deliberative citizenship makes or would make demands on the capacities of citizens which are unrealistic. The idea that deliberating citizens will be able to make political judgments without considering their personal preferences and the idea that they have a capacity to critically reflect on and perhaps revise their existing motives, beliefs, and preferences are certainly demanding ideas. While I admit that it may be that citizens are *unlikely* to develop and utilize that reflective capacity at all times during their lives, I maintain that the capacity is within the limits of practical possibility and I offer three reasons to support the claim that the conception of deliberative citizenship relied upon here is not unrealistic:

1. The Deep Structure of Human Psychology

The first reason has to do with human nature and psychology itself. Though citizens often act from self-interested motivations or perhaps with a lack of awareness of the consequences of their actions, nothing in the structure of human nature itself rules out the possibility that human beings can reflect on and perhaps revise many of their interests, preferences, motives, and beliefs. As Rousseau discovered, there is a distinction to be made between human nature – which is naturally good and unalterable – and the determinate social ethos which is rigid, but nevertheless malleable. When in political theory we take people “as they are”, we recognize the limits imposed by the deep structure of human nature but adopt a critical attitude towards determinate preferences, interests, motives, and beliefs which are shaped in

social life (Rousseau, 1978: 1.1).¹²⁴ The idea of deliberative citizenship holds that citizens have the *capacity* to distinguish between their constructed preferences, interests, motives, and beliefs, on the one hand, and the fixed features of human nature on the other. Even if that capacity is rarely, if ever, exercised, its existence means that we are not being wildly utopian in holding that citizens can reflect on and perhaps revise their views.

2. The Dynamic Character of Comprehensive Doctrines

Second, a feature of the structure and content of most comprehensive doctrines also points towards the realism of the conception of the citizen relied upon here. Citizens rarely have worldviews which are perfectly settled, coherent, and comprehensive so there is a certain amount of room to maneuver within their comprehensive doctrines. Writing about cultural worldviews, Benhabib notes that much contemporary theorizing adopts “faulty epistemic premises” about cultural rigidity (2002: 4). In particular she holds that cultures are assumed by many theorists to be much more homogenous, bounded, and distinct and that their content is much more fixed than is, in fact, the case (2002:4). By contrast, Benhabib adopts a social constructivist view about culture which holds that the content of cultural worldviews is the product of ongoing evaluations, narratives, and actions (2002:5). That is, because the “social observer” studies cultural worldviews from an external standpoint, the “social observer” often fails to recognize the “contested and contestable” character of cultural worldviews as they are experienced by “participants in the culture” (2002: 5). Thus, she continues, “we should view human cultures as constant creations, recreations, and negotiations” and recognize that “existing social and cultural cleavages are transformed into political mobilization” (2002: 8, 17). Insofar as ethical, religious, and other *non-cultural*

¹²⁴ On the distinction between human nature and social ethos, see Rousseau (1986).

comprehensive worldviews are characterized by this same contestability, there is room *within* comprehensive doctrines for maneuver which contributes to the realism of the idea of deliberative citizenship.¹²⁵

When asked whether they could endorse a conception of justice and an idea of deliberative citizenship from within their own worldview, then, most citizens are able to consult a worldview which has a certain elasticity and therefore may be able to locate support for the conception of justice and the idea of deliberative citizenship by re-ordering or revising certain elements of their worldview. While some critics might regard this as a demand that citizens conform to an alien conception of justice, on a Rawlsian understanding this is simply to ask citizens to locate support within their comprehensive views for a conception of justice and citizenship which they already find attractive for other, non-comprehensive, reasons.

In a similar manner, David Miller notes that empirical research about what ordinary people think about justice “suggests not that people are locked into fixed beliefs about justice, but that there is a very considerable degree of indeterminacy in many of their specific beliefs” (1999: 60). They may have certain background commitments to *general ideas* about fairness, equality, and the like, but are rarely “locked into” specific beliefs about justice (1999: 60).¹²⁶ So the dynamic rather than static, and constructed rather than primordial, nature of most worldviews is another reason why a conception of the citizen as having a capacity to reflect on and perhaps to revise their interests and beliefs is not unrealistic.¹²⁷

¹²⁵ Benhabib is careful to note that “sociological constructivism does not suggest that cultural differences are shallow or somehow unreal or ‘fictional’. Cultural differences run very deep and are very real....Yet, the student of human affairs should never take groups’ and individuals’ cultural narratives at face value. Rather, to explain human behaviour, the student should seek to understand the totality of circumstances of which culture is an aspect” (2002: 7).

¹²⁶ See also Miller (1999: 61-92).

¹²⁷ See also Rawls (2001: 193) on the “looseness” of comprehensive views.

3. The Complementarity of Institutions and Deliberative Citizenship

Finally, though asking citizens to reflect on their determinate interests and beliefs as individuals may be a demanding request, such reflection can be encouraged and assisted when deliberative institutions and opportunities are available in the formal and the informal public spheres. Institutional arrangements which encourage democratic deliberation, I argue, can help citizens to be more reflective about their determinate preferences, opinions, and beliefs and help them to acquire the dispositions and developed capacities necessary for deliberative citizens to discover and to support a conception of justice. And evidence from a few practical experiments in deliberative democracy indicates that those expectations are not ill-founded.¹²⁸ Evidence of the effects that deliberative engagements with other citizens have on individuals' interests, beliefs, and motives, then, offers another reason why a conception of the citizen as having a capacity to reflect on and perhaps to revise interests, preferences, motives, and beliefs is not unrealistic.

However, the idea that institutions can facilitate the exercise of deliberative citizenship seems to be in tension with the earlier claim that deliberative citizenship must precede institutions. The dynamic model of institutions and deliberative citizenship might strike some as itself objectionably utopian or even tautological. The institutions which are to encourage deliberative citizenship seem to depend for their existence on the dispositions and activities of deliberative citizens. But that view is not, I think, objectionably utopian nor tautological once we recognize that neither completely just institutions nor perfect deliberative citizenship are necessary for our solution to the problem of practical possibility.

¹²⁸ Recent studies on participatory-budgeting in Porto Alegre, Brazil demonstrate that deliberation can indeed encourage more reflective and "enlarged" thinking among ordinary citizens. See, for

The ideal of deliberative citizenship, while demanding and often in need of institutional encouragement, is a reflective attitude that citizens are capable of adopting whether institutions to encourage it exist or not. When I say that deliberative citizenship can be regarded as more realistic given that certain institutional arrangements can facilitate or encourage it, that does not mean that deliberative citizenship is impossible without those institutions. Absent those institutions, deliberative citizenship is admittedly improbable, but not impossible; whereas with those institutions, deliberative citizenship is both possible and less improbable.

Of course, that will fail to satisfy those realists who want a closer fit between justice and realism. But as I have argued throughout the dissertation, a closer fit between justice and realism would risk adjusting the content of justice to accommodate objectionable features of human beings and the social and political status quo. Solving the problem of practical *possibility*, rather than aiming at guarantees of motivational stability and doctrinal compatibility, is a more appropriate resolution to the tension between justice and realism.

V. Justice, Legitimacy, and Realism

Whether or not the conception of deliberative citizenship is realistic and whether or not it helps to solve the tension between justice and realism aside, there is the large, still unanswered, utopian challenge about why we should care about the demands of realism at all: Why should we regard motivational possibility and doctrinal compatibility as matters of justice and not simply matters of practical application?

The idea of normatively salient demands of realism in the first part of this chapter goes some way towards responding to that challenge. Some existing features of human beings

example, Abers (1999) as well as the essays by Abers and Baiocchi in Fung & Wright (2003). For

and the world, I said, are not mere practical constraints on a conception of justice but rather fundamental interests of democratic citizens that must be accommodated if a conception of justice is to be considered one to which reasonable citizens might agree. Reasonable pluralism and a concern for liberty, I argued, would have to be accommodated in a conception of justice if that conception is to be considered attractive for a democratic society.

Moreover, because the problem of realism is treated here more as a matter of demonstrating practical *possibility* than one of accommodating all existing motives and beliefs, the utopian concern that principles of justice will make concessions to objectionable motives and beliefs in order to achieve stability is unfounded. The approach to the demands of realism defended here shows how the ideals of deliberative democracy and deliberative citizenship, although demanding, are not unrealistic. Because the solution to the tension between justice and realism appeals to ideals or conceptions, rather than citizens' existing motives, beliefs, and the imperfect practice of democratic deliberation, the utopian need not worry about concessions to objectionable features of human beings and political practice. Indeed, it is hard to see why citizens would agree to principles of justice which are adjusted to account for objectionable, but alterable, features of the social and political status quo, rather than principles which maintain a critical attitude and emancipatory hope in the face of an objectionable status quo.

In my view, agreement on principles of justice by reasonable citizens will only be achieved when it can be demonstrated that those principles are motivationally possible and capable of endorsement from within reasonable comprehensive doctrines. Just as it would be unreasonable to expect all citizens to agree to principles of justice which would regard

some cautionary remarks, see Mendelberg (2001).

significant inequality or oppression as just, it would be unreasonable to expect citizens to agree to principles of justice which demand the impossible. And by impossible, I don't simply mean impossible in the face of existing motives, beliefs, interests, and behaviour, but impossible in the face of deep, unalterable, structural features of human psychology.

In the Introduction I said that, on my view, a conception of justice must be a conception to which citizens agree or could agree. But I suggested that, even if we can locate that agreement, once the conception of justice is in place, it is likely that some citizens will violate or be tempted to violate the demands of justice from time to time – even when those citizens have, or could have, endorsed the principles which they subsequently violate. In that case, the use or threatened use of coercive measures to ensure compliance with the demands of justice may be legitimate. But in the face of that possible coercion, I maintain that no one could rationally or reasonably agree to principles of justice which are either motivationally impossible or with which reasonable comprehensive doctrines are incompatible because doing so would be to agree to be subjected to inevitable and futile coercion. That is, if the demands of justice are beyond the motivational capacities of human beings or require commitments and behaviour which reasonable citizens attached to reasonable comprehensive doctrines could not fulfill, then citizens would inevitably face coercive action and, moreover, because changes in behaviour or attitude are either motivationally or doctrinally impossible, the coercion would serve no practical purpose. Thus, I conclude that because no rational or reasonable citizen would agree to unavoidable and pointless coercive action, only a conception of justice which offered a solution to the problems of motivational possibility and doctrinal compatibility could achieve legitimacy.

That it will be *difficult* to realize demanding principles might prompt citizens to take steps to protect themselves against exploitation in actual political life. In that sense the

assurance problem is a problem of application, but not one of justice – that is, while it would be prudent to take steps to ensure that the motivated will not be exploited by the unmotivated (or the moral by the immoral), for example, the content of those measures should not affect the content of a defensible conception of justice. If such concessions are impermissible then the principles of justice will remain demanding. But if those principles do not make demands that would break the limits of motivational possibility or doctrinal compatibility, then citizens might still agree that those principles best capture the conception of justice which should guide public political deliberation.

Nevertheless, while the idea of normatively salient demands of realism and the ideals (rather than practice) of deliberative democracy and deliberative citizenship go some way towards responding to the utopian objection, there is a deeper utopian objection to which we must respond. Utopians might ask why we should be concerned with agreement and legitimacy in constructing principles of justice at all. Indeed, G.A. Cohen (2003) asks what legitimacy or the idea of agreement really add to the justness of principles of justice.

Agreement or legitimacy are necessary because that is one way in which a conception of justice can recognize and embody the idea that citizens ought to be regarded as free and equal members of a political community. While principles of justice might appear to protect citizens' freedom within a political sphere already regulated by the conception of justice, unless those citizens could be regarded as having actively agreed to that initial conception of justice, it is hard to see how we have really recognized and respected the fact that they are to be treated as politically free. If one thinks that justice is something that exists prior to the experiences and activities of citizens in social and political life then one will likely regard legitimacy and agreement as superfluous. But if one thinks that justice is something shaped and reshaped in reflective equilibrium with the conditions of our social and political lives as

we find them, then legitimacy and agreement will matter to the justness of a conception of justice.

VI. Conclusion: From Resignation to Hope

Though we might hold that the ideals of deliberative democracy and deliberative citizenship are not *unrealistic*, it might still be thought *unlikely* that citizens will usually, or even occasionally, adopt the reflective attitude of deliberative citizens necessary to solve the problem of practical possibility. Given the conflict between what one might regard as rational in light of one's determinate interests, preferences, and beliefs and what would be regarded as reasonable, it might be too much to expect citizens to choose the reasonable over the rational when the two seem to conflict. But while the conception of deliberative citizenship might for that reason seem utopian, that does not make it pejoratively so. A realistically utopian conception of justice should respect the limits of what is practically possible, but it should maintain a critical attitude towards objectionable motives, beliefs, interests, preferences, and behaviour which is commonplace but not unalterable.

The issue, to repeat, is to consider just what role the demands of realism should play in theorizing about justice. I want to conclude by highlighting three distinctions that, in one way or another, have weaved their way through the dissertation and that I maintain will help us answer that concern about the appropriate role of the demands of realism.

The first is a distinction between practical demands which are fixed and practical demands which are amenable to transformation. A conception of justice should respect the limits imposed by fixed constraints but it should not resign itself to those unattractive features of the world and determinate motives, interests, preferences, and beliefs which are not part of the deep structure of human nature. In the dissertation I have defended the view

that certain demands – such as a concern for freedom and a desire to pursue a reasonable conception of the good – should be acknowledged and accommodated by a conception of justice. But justice should not resign itself to the content of unreasonable worldviews – i.e., views held by citizens who insist on the exclusive truth of their views and who reject the need to offer reasons for their claims that other reasonable citizens could endorse – or to unattractive and objectionable motives and interests. If such accommodations were even conceptually possible under conditions of ethical and religious pluralism, we would still have no reason to capitulate to the demands of unreasonable worldviews because that would force us to relinquish the emancipatory promise of justice.

The second distinction that has left its mark here is a distinction between principles of justice on the one hand and the *application* of those principles in real political life, on the other. Though we should strive to realize the ideals of an attractive conception of justice when applying justice to existing societies, we should be sensitive to the practical constraints faced by contemporary democracies and adjust our expectations of feasible application as required. Indeed, it would be highly desirable to provide guarantees to those who are determined to be just that their behaviour will not be exploited by citizens who are out to advance their own interests regardless of the effects on others. If such guarantees cannot be provided, it may be prudent to release otherwise justly motivated citizens from those requirements of justice which require significant self-sacrifice and expose them to exploitation by the unjust. But releasing citizens from those demands for practical or prudential reasons in the application of principles of justice does not mean that the content of justice itself should be altered. Surely we can distinguish between fair and unfair demands in practice without giving up on the idea that the demands of justice should retain their critical and emancipatory content. To be sure, the fact that we can even speak of *releasing*

citizens from the obligations of justice under certain circumstances implies that the content of justice itself remains unchanged.

Thus, political theory should refrain from accommodating the demands of realism in the construction of principles of justice except in those cases where the demands require accommodation as a matter of justice itself. In the dissertation I have defended the view that reasonable pluralism and normatively salient motives are such demands whereas unreasonable pluralism or the self-interested motivation which produces the assurance problem and the desire for relative superiority over others, for example, are not. A conception of justice must accommodate the range of reasonable pluralism and normatively salient motives because they would emerge under otherwise well-ordered and just political arrangements, they are held by citizens who themselves as reasonable, and because they constitute interests which citizens would have *as democratic citizens* and not merely as unreflective strategic actors.

The final distinction that has played a guiding role in the dissertation is a distinction between political optimism and pessimism, on the one hand, and political hope and despair, on the other. I defined political optimism in the introduction as the expectation that the society in which one lives will become more just over time. And I defined political pessimism, by contrast, as the view which holds that a society not only will not become more just over time, but as the expectation that it may even deteriorate from the perspective of justice. Yet, I have argued that we should not look to the ideas of political optimism and pessimism to give us the parameters within which we should theorize about justice. Instead, I argued that we should be guided by an idea of political hope. Political hope is the view that a better world is possible, even if unlikely. It holds that the limits of human capacity have not yet been reached and that by furthering developing and exercising those capacities –

especially the capacity to reflect on one's own motives, preferences, and interests and to change those when doing so would create a more just world – the possibility of a just world is within human reach. As Kant puts it, “man possesses a greater moral capacity, still dormant at present, to overcome eventually the evil principle within him (for he cannot deny that it exists), and to hope that others will do likewise” (1970b: 103). While we need not endorse Kant's conviction that there is an evil principle in human nature,¹²⁹ the hope he expresses for humanity is one that we should adopt as a guide in political theory.

Though some may regard the attempt to accommodate only normatively salient demands of realism and the reliance on a conception of deliberative citizenship to solve the problem of practical possibility as evidence of a naïve political optimism, we should instead regard these things as an attempt to construct an attractive conception of justice rooted in the hope that such a conception is simply possible. It is rooted in the hope that it is possible to sever the Thrasymachian link between power and justice. As Rawls puts it, “political philosophy is realistically utopian when it extends what are *ordinarily thought* to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition (1999b: 11. My emphasis). In short, a realistically utopian approach to justice limits its hopes only by what is in fact practically possible, not by what is politically probable nor by what is given by the political status quo. If theorists of justice take their cues from an unjust political status quo, and regard that status quo as unalterable, then it is hard to see how theorizing about justice could produce anything more than a political theory of despair.

Thus, a realistically utopian approach to principles of justice does not assume that all features of the world – that all human motives as they are currently found – are fixed. Indeed, in thinking about justice and the demands motivational and doctrinal compliance, as

¹²⁹ On this conviction, however, see Kant (1960).

well as the demands of realism more generally, we would do well to remind ourselves of Beitz's thoughts about the possible and the impossible. While Beitz agrees that justice should not demand the impossible in the sense of demanding that justice ignore or seek to overcome "impediments that are unalterable or unavoidable", he maintains that "impediments to change that are themselves capable of modification over time" should not be included in our notion of the impossible (1999: 156). Similarly, I have argued that we should not take as the limits of the possible the existing motives and beliefs of human beings; rather the limits of possibility should be fixed only by what we might hope for.

And what might we hope for? Given the limits of human capacity, we might hope for a conception of justice which includes protections of citizens' freedom and equality as democratic citizens as well the rights and resources necessary for citizens to pursue reasonable conceptions of the good. Thus, we need not resign ourselves to the current array of objectionable preferences, interests, beliefs, and motives of human beings nor, for that matter to power inequalities and other unattractive features of the social and political status quo. Of course, the fact that we may hope for a conception of justice which is not corrupted by objectionable features of human beings and the social and political status quo does not mean that we can be any less pessimistic or any more optimistic about the emergence of a just world than we might already have been. But it does permit us to avoid the politics of despair. And if Kant is right that when justice perishes, "there is no longer any value in human beings' living on earth" (1996: 105) then avoiding the politics of despair and pursuing a politics of hope might be our best chance for survival and, perhaps, even our best chance for happiness.

Bibliography

- Abers, R. (1999). Practicing Radical Democracy: Lessons from Brazil. *Plurimondi*, 1 (2), 67-82.
- An-Na'im, A. A. (1990). *Toward and Islamic Reformation: Civil Liberties, Human Rights, and International Law*. Syracuse: Syracuse University Press.
- An-Na'im, A. A. (1992). Toward a Cross-Cultural Approach to Defining International Standards of Human Rights: The Meaning of Cruel, Inhuman, or Degrading Treatment or Punishment. In A.A. An-Na'im (Ed.), *Human Rights in Cross-Cultural Perspectives: A Quest for Consensus*. Philadelphia: University of Pennsylvania Press.
- Aristotle. (1958). *The Politics*. London: Oxford University Press.
- Baier, K. (1999). Justice and the Aims of Political Philosophy. *Ethics*, 99, 771-790.
- Baynes, K. (1992). *The Normative Grounds of Social Criticism: Kant Rawls, and Habermas*. Albany: State University of New York Press.
- Beitz, C. (1999). *Political Theory and International Relations*. Princeton: Princeton University Press.
- Benhabib, S. (1986). *Critique, Norm, and Utopia*. New York: Columbia University Press.
- Benhabib, S. (1992). *Situating the Self*. New York: Routledge.
- Benhabib, S. (2002). *The Claims of Culture*. Princeton: Princeton University Press.
- Benhabib, S. (Ed.). (1996). *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton: Princeton University Press.
- Benhabib, S. & Cornell, D. (Eds.). (1987). *Feminism as Critique: Essays on the Politics of Gender in Late-Capitalist Societies*. Minneapolis: University of Minnesota Press.
- Benhabib, S. & Dallmayr, F. (Eds.). (1990). *The Communicative Ethics Controversy*. Cambridge: MIT Press.
- Berlin, I. (1969). *Four Essays on Liberty*. Oxford: Oxford University Press.
- Blaug, R. (2000). Citizenship and Political Judgment: Between Discourse Ethics and *Phronesis*. *Res Publica* 6, 178-198.
- Bohman, J. & Rehg, W. (Eds.). (1997). *Deliberative Democracy: Essays on Reason and Politics*. Cambridge: MIT Press.
- Bourdieu, P. (1998). *Practical Reason*. Stanford: Stanford University Press.

- Bufacchi, V. (2005). Motivating Justice. *Contemporary Political Theory*, 4, 25-41.
- Calhoun, C. (Ed.). (1992). *Habermas and the Public Sphere*. Cambridge: The MIT Press.
- Camus, A. (1991). The Myth of Sisyphus. In A. Camus, *The Myth of Sisyphus and other essays*. New York: Vintage International.
- Carr, E.H. (1964). *The Twenty Years' Crisis, 1919 – 1939*. New York: Harper & Row.
- Cassirer, E. (1954). *The Question of Jean-Jacques Rousseau*. New York: Columbia University Press.
- Chambers, S. (1996). *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse*. Ithaca: Cornell University Press.
- Cohen, G. A. (2000). *If You're an Egalitarian, How Come You're So Rich?* Cambridge: Harvard University Press.
- Cohen, G. A. (2003). Rescuing Justice From Constructivism. (Unpublished MS).
- Cohen, G.A. (2003b). Facts and Principles. *Philosophy and Public Affairs* 31: 3 (Summer).
- Cohen, J. (1993). Moral pluralism and political consensus. In D. Copp, J. Hampton & J. Roemer, (eds.). *The Idea of Democracy*. Cambridge: Cambridge University Press.
- Cohen, J. (1996). Procedure and Substance in Deliberative Democracy. In S. Benhabib (ed). *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton: Princeton University Press.
- Cohen, J. (1997). Deliberation and Democratic Legitimacy. In J. Bohman & W. Rehg (Eds.). *Deliberative Democracy: Essays on Reason and Politics*. Cambridge: MIT Press.
- Cohen, J. (1999). Reflections on Habermas and Democracy. *Ratio Juris*, 12 (4), 385-416.
- Cohen, J. (2001). A Free Community of Equals. (Unpublished MS).
- Cohen, J. (2002) Taking People as They Are? *Philosophy & Public Affairs* 30 (4), 363-386
- Cohen, J. (2003). For a Democratic Society. In S. Freeman (ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.
- Cohen, J. & Fung, A. (1996). Introduction: Just Institutions. In *Constitution, Democracy, and State Power: The Institutions of Justice*. Cheltenham: Edward Elgar.
- Cohen, J. & Rogers, J. (1995). Secondary Associations and Democratic Governance. In Wright, E.O. (Ed.). *Associations and Democracy*. New York: Verso.

- Cronin, C & De Greiff, P. (1998). Introduction. In J. Habermas, *The Inclusion of the Other*. C. Cronin & P. De Greiff (Eds.). Cambridge, MIT Press.
- Deveaux, M. (2000). *Cultural Pluralism and Dilemmas of Justice*. Ithaca: Cornell University Press.
- Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge: Harvard University Press.
- Elster, J. (1989). *The Cement of Society: A Study of Social Order*. Cambridge: Cambridge University Press.
- Elster, J. (Ed.). (1998). *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Estlund, D. (1998). Liberalism, Equality, and Fraternity in Cohen's Critique of Rawls. *The Journal of Political Philosophy*, 6 (1), 99-112.
- Farrelly, C. (2004). *An Introduction to Contemporary Political Theory*. London: Sage Publications.
- Freeman, S. (1990). Reason and Agreement in Social Contract Views. *Philosophy and Public Affairs*, 19 (2), 122-157.
- Freeman, S. (2000). Deliberative Democracy: A Sympathetic Comment. *Philosophy and Public Affairs*, 29 (4), 371-418.
- Freeman, S. (Ed.). (2003). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.
- Fung, A. & Wright, E.O. (Eds.). (2003). *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. London: Verso Press.
- Galston, W. (1989). Pluralism and Social Unity. *Ethics*, 99, 711-726.
- Gauthier, D. (1986). *Morals By Agreement*. Oxford: Clarendon Press.
- Gutmann, A. & Thompson, D. (1996). *Democracy and Disagreement*. Cambridge: Belknap.
- Gutmann, A & Thompson, D. (2004). *Why Deliberative Democracy?* Princeton: Princeton University Press.
- Habermas, J. (1984). *The Theory of Communicative Action*. Vol. 1. Thomas McCarthy, trans.. Boston: Beacon Press.
- Habermas, J. (1987). *The Theory of Communicative Action*. Vol. 2. Thomas McCarthy, trans.. Boston: Beacon Press.
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere*. T. Burger (trans.). Cambridge: MIT Press.

- Habermas, J. (1990). *Moral Consciousness and Communicative Action*. C. Lenhardt & S. Weber NicholSEN, (trans.). Cambridge: MIT Press.
- Habermas, J. (1993). *Justification and Application: Remarks on Discourse Ethics*. Cambridge: MIT Press.
- Habermas, J. (1996). *Between Facts and Norms*. W. Rehg (trans.). Cambridge: MIT Press.
- Habermas, J. (1998). *The Inclusion of the Other*. C. Cronin & P. De Greiff (Eds.). Cambridge, MIT Press.
- Hamilton, A., Madison, J. & Jay, J. (1982). *The Federalist Papers*. G. Willis (Ed.). New York: Bantam Books.
- Hand, L. (1952). *The Spirit of Liberty*. New York: Knopf.
- Heath, J. (2001). *Communicative Action and Rational Choice*. Cambridge: MIT Press.
- Heath, J. (2002). *The Efficient Society: Why Canada is as Close to Utopia as it Gets*. Toronto: Penguin Canada.
- Heath, J. (2005). Rawls on Global Distributive Justice: A Defense. Forthcoming in D. Weinstock (Ed.). *Canadian Journal of Philosophy Supplementary Volume 29*, Lethbridge: University of Calgary Press.
- Heath, J. (2005b). On the Scope of Egalitarian Justice. (Unpublished MS).
- Hegel, G. W. F.. (1967). *Philosophy of Right*. T. M. Knox (trans.). Oxford: Oxford University Press.
- Hobbes, T. (1968). *Leviathan*. C.B. Macpherson (Ed.). London: Penguin Books.
- Hobbes, T. (1991). *Leviathan*. R. Tuck (Ed.). Cambridge: Cambridge University Press.
- Holmes, S. (1988). Precommitment and the paradox of democracy. In J. Elster & R. Slagstad (Eds.). *Constitutionalism and Democracy*. Cambridge: Cambridge University Press.
- Hume, D. (1978). *A Treatise of Human Nature*. L. A. Selby-Bigge (Ed.). Oxford: Clarendon Press.
- Hume, D. (1983). *An Enquiry Concerning the Principles of Morals*. J. B. Schneewind (Ed.). Indianapolis: Hackett.
- Hume, D. (1985). Of the Origin of Government. In D. Hume, *Essays: Moral, Political, and Literary*. E.F. Miller (Ed.). Indianapolis: Liberty Classics.

- Kant, I. (1960). *Religion Within the Limits of Reason Alone*. T. M. Greene & H. H. Hudson (trans.). New York: Harper & Row.
- Kant, I. (1970a). On the Common Saying: 'This May be True in Theory, but it does not Apply in Practice'. In I. Kant, *Political Writings*. H. Reiss (Ed.). Cambridge: Cambridge University Press.
- Kant, I. (1970b). Perpetual Peace: A Philosophical Sketch. In I. Kant, *Political Writings*. H. Reiss (Ed.). Cambridge: Cambridge University Press.
- Kant, I. (1990a). *Foundations of the Metaphysics of Morals*. L. W. Beck (trans.). New York: MacMillan.
- Kant, I. (1990b). *Critique of Pure Reason*. J. M. D. Meiklejohn (trans.). New York: Prometheus Books.
- Kant, I. (1993). *Critique of Practical Reason*. L. W. Beck (trans.). New Jersey: Prentice Hall.
- Kant, I. (1996). *The Metaphysics of Morals*. M. Gregor (Ed.). Cambridge: Cambridge University Press.
- Kymlicka, W. (1989). Liberal Individualism and Liberal Neutrality. *Ethics*, 99, 883-905.
- Larmore, C. (2003). Public Reason. In S. Freeman (Ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.
- Lear, J. (1998). *Open-Minded: Working Out the Logic of the Soul*. Cambridge: Harvard University Press.
- Lemos, R. M. (1977). *Rousseau's Political Philosophy: An Exposition and Interpretation*. Athens: University of Georgia Press.
- Locke, J. (1980). *Second Treatise of Government*. C.B. Macpherson (Ed.). Indianapolis: Hackett.
- Machiavelli, N. (1979). *The Portable Machiavelli*. P. Bondanella & M. Musa (Eds. and trans.). New York: Penguin.
- Manin, B. (1987). On Legitimacy and Political Deliberation. *Political Theory*, 15, 3.
- Masters, R. D. (1968). *The Political Philosophy of Rousseau*. Princeton: Princeton University Press.
- McCarthy, T. (1990). Introduction. In J. Habermas. *Moral Consciousness and Communicative Action*. C. Lenhardt & S. Weber Nicholsen, (trans.). Cambridge: MIT Press.

- Mendelberg, T. (2001). The Deliberative Citizen: Theory and Evidence. Forthcoming in M. D. Carpin, L. Huddy & R. Y. Shapiro (Eds.). *Research in Micropolitics, Volume 6*. Elsevier Press.
- Mill, J.S. (1991). Utilitarianism. In J.S. Mill, *On Liberty and Other Essays*. J. Gray (Ed.). Oxford: Oxford University Press.
- Miller, D. (1999). *Principles of Social Justice*. Cambridge: Harvard University Press.
- Murphy, L. B. (1999). Institutions and the Demands of Justice. *Philosophy & Public Affairs*, 27 (4), 251-291.
- Niebuhr, R. (1944). *Children of Light and the Children of Darkness*. Charles Scribner's Sons. Retrieved June 30, 2005, from the Religion-Online website: <http://www.religion-online.org/showbook.asp?title=446>.
- Okin, S. (1989). *Justice, Gender, and the Family*. Basic Books.
- Okin, S. (1999). Is Multiculturalism Bad for Women? In J. Cohen, M. Howard & M. C. Nussbaum (Eds.). *Is Multiculturalism Bad for Women?* Princeton: Princeton University Press.
- Okin, S. (2005). Multiculturalism and Feminism: No Simple Question, No Simple Answers. Forthcoming in A. Eisenberg & J. Spinner-Havel (Eds.). *Minorities within Minorities*. Cambridge: Cambridge University Press.
- Pitkin, H. (1981). Justice: On Relating Private and Public. *Political Theory*, 9 (3), 327-352.
- Plato. (1974). *Republic*. G. M. A. Grube (Trans.). Indianapolis: Hackett.
- Plato. (1988). *Laws*. T. Pangle (Trans.). Chicago: University of Chicago Press.
- Pogge, T. (2000). On the Site of Distributive Justice: Reflections on Cohen and Murphy. *Philosophy & Public Affairs*, 29 (2), 137-169.
- Popper, K. (1966). *The Open Society and Its Enemies, Volume 1: The Spell of Plato*. Princeton: Princeton University Press.
- Rawls, J. (1996). *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. (1999). *A Theory of Justice*. Revised edition. Cambridge: Harvard University Press.
- Rawls, J. (1999b). *The Law of Peoples*. Cambridge: Harvard University Press.
- Rawls, J. (1999c). The Idea of Public Reason Revisited. In J. Rawls, *The Law of Peoples*. Cambridge: Harvard University Press.
- Rawls, J. (1999d). *Collected Papers*. S. Freeman (Ed.). Cambridge: Harvard University Press.

- Rawls, J. (2000). *Lectures on the History of Moral Philosophy*. B. Herman (Ed.). Cambridge: Harvard University Press.
- Rawls, J. (2001). *Justice as Fairness: A Restatement*. E. Kelly (Ed.). Cambridge: Belknap Press of Harvard University Press.
- Rehg, W. (1996). Translator's Introduction. In J. Habermas. *Between Facts and Norms*. W. Regh (Trans.). Cambridge: MIT Press.
- Rosenblatt, H. (1997). *Rousseau and Geneva*. Cambridge: Cambridge University Press.
- Rostbøll, C. F. (2005). Preferences and Paternalism: On Freedom and Deliberative Democracy. *Political Theory*, 33 (3), 370-396.
- Rousseau, J.J. (1964). *The First and Second Discourses*. R. D. Masters (Ed. and trans.) & J. R. Masters (trans.). New York: St. Martin's Press.
- Rousseau, J.J. (1968). *Letter to d'Alembert on the Theatre*. A. Bloom (Trans.). Ithaca: Cornell University Press.
- Rousseau, J.J. (1978). *On the Social Contract with Geneva Manuscript and Political Economy*. R. D. Masters (Ed.) & J.R. Masters (Trans.). New York: St. Martin's Press.
- Rousseau, J.J. (1985). *The Government of Poland*. W. Kendall (Trans.). Indianapolis: Hackett.
- Rousseau, J.J. (1986).. *First and Second Discourses and Essay on the Origin of Languages*. V. Gourevitch (Ed. & trans.). New York: Harper & Row.
- Rousseau, J.J. (1987). Discourse on the Origin and Foundations of Inequality Among Men. In J.J. Rousseau, *The Basic Political Writings*. D.A. Cress (Trans.). Indianapolis: Hackett Publishing Company.
- Rousseau, J.J. (1993). *Emile*. B. Foxley (Trans.). London: Everyman.
- Sandel, M. (1982). *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Saramago, J. (1997). *Blindness*. San Diego: Harcourt.
- Scanlon, T.M.. (1998). *What We Owe to Each Other*. Cambridge: Belknap.
- Scheuerman, W. E. (1999). Between Radicalism and Resignation: Democratic Theory in Habermas's *Between Facts and Norms*. In P. Dews (Ed.). *Habermas: A Critical Reader*. Oxford: Blackwell Publishers.
- Shapiro, I. (2003). *The State of Democratic Theory*. Princeton: Princeton University Press.

- Shklar, J. (1969). *Men and Citizens: A Study of Rousseau's Social Theory*. Cambridge: Cambridge University Press.
- Talmon, J.L. (1952). *The Rise of Totalitarian Democracy*. Boston: The Beacon Press.
- Vernon, R. (2001). *Political Morality: A Theory of Liberal Democracy*. London: Continuum.
- White, S. (2003). *The Civic Minimum: On the Rights and Obligations of Economic Citizenship*. Oxford: Oxford University Press.
- Williams, A. (1998). Incentives, Inequality, and Publicity. *Philosophy & Public Affairs*, 27 (3), 225-247.
- Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
- Wokler, R. (1995). *Rousseau*. Oxford: Oxford University Press.
- Wolff, J. (1998). Fairness, Respect, and the Egalitarian Ethos. *Philosophy & Public Affairs*, 27 (2), 97-122.