

John Rawls's Actual Contractualism

William Thomas M Porter

University College London

PhD Degree in Philosophy

UMI Number: U592602

All rights reserved

INFORMATION TO ALL USERS

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

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

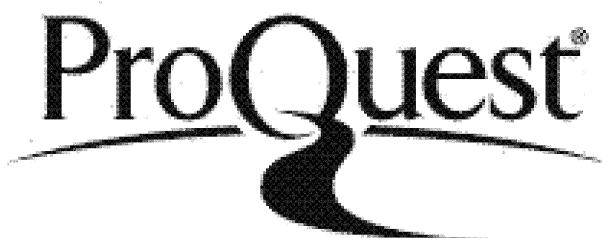


UMI U592602

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



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

Abstract

This thesis argues for an unorthodox interpretation of John Rawls's egalitarianism as a hybrid of 'actual contractualism' and 'modal contractualism'. It also offers a defence of the theory so understood. According to actual contractualism, a system of political institutions and norms is just only if each person over whom it claims authority actually accepts it in some sense. Actual contractualists stand in contrast with modal contractualists, who take justice to require that no one could reasonably reject the institutions and norms in question. Rawls is standardly read as a modal contractualist, but I argue that his view includes an significant element of actual contractualism.

The thesis is divided into three parts. The first part describes actual contractualism and contrasts it with modal contractualism. It goes on to consider the possibility of a hybrid theory, which appeals to modal contractualist reasons to justify an actual contractualist test for justice. I suggest that this is an attractive view. In the second part I go on to argue that a careful understanding of Rawls's theory view reveals it to be hybrid contractualist. I elaborate Rawls's strategy of 'political constructivism' in the light of this interpretation, and attempt to show that it is very much in the Lockean actual contractualist tradition.

The final part of the thesis concerns the justification of specifically Rawls's egalitarianism. I contend that Rawls's argument for justice as fairness can be seen as a detailed effort to explain why his egalitarianism is, in the relevant sense, actually accepted by each person, and I argue that as such it succeeds. I then contrast the Rawlsian view with left-libertarianism, another attempt to marry actual contractualism and egalitarianism. I argue that Rawls's is the more thoroughgoing, unified view, and should be preferred on that basis.

Table of Contents

Acknowledgments	6
Introduction	
1. <i>Contractualism and the individualistic ideal of justification</i>	8
2. <i>Rawls and hybrid contractualism</i>	14
3. <i>Hybrid contractualism and egalitarianism</i>	18
PART I: CONTRACTUALIST FRAMEWORKS	23

Chapter 1: Actual Contractualism

1. <i>Actual contractualism</i>	24
2. <i>Actual contractualism and approaches to justice</i>	26
3. <i>Two variables: scope and acceptance</i>	35
4. <i>The shape of actual contractualism</i>	39
5. <i>Hobbes, Locke, Nozick, and actual contractualism</i>	48
6. <i>Autonomy and the appeal of actual contractualism</i>	62

Chapter 2: Modal Contractualism and Hybrid Contractualism

1. <i>Modal contractualism</i>	68
2. <i>What's distinctive about modal contractualism? – the question</i>	71
3. <i>What's distinctive about modal contractualism? – the answers</i>	75
4. <i>First-order normative modal contractualism</i>	89
5. <i>Modal contractualism and actual contractualism</i>	94
6. <i>Two puzzles for hybrid theorists</i>	99
7. <i>Conclusion</i>	108

PART II: RAWLS AND HYBRID CONTRACTUALISM 111

Chapter 3: Rawlsian Contractualism

1. <i>Introduction</i>	112
2. <i>Political Liberalism and hybrid contractualism</i>	114
3. <i>The burdens of judgment as a basis for actual contractualism</i>	127
4. <i>The hybrid view and the actual contractualist motivation</i>	140
5. <i>The hybrid view and acceptance</i>	152

Chapter 4: Political Constructivism and the Hybrid View

1. <i>Introduction</i>	161
2. <i>The problem of practical conflict</i>	165
3. <i>Shared premises and the priority of justice</i>	168
4. <i>Indoctrination</i>	180
5. <i>False beliefs and loose beliefs</i>	182
6. <i>Mistaken interests</i>	187
7. <i>Wrong background</i>	195

PART III: HYBRID CONTRACTUALISM AND EGALITARIANISM 210

Chapter 5: Hybrid Contractualism and Egalitarianism

1. <i>Introduction</i>	211
2. <i>The informal argument, stage 1: arbitrariness and the erosion of prior claims</i>	215
3. <i>The informal argument, stage 2: equality and the significance of cooperation</i>	223
4. <i>The informal argument, stage 3: from equality to the difference principle</i>	228
5. <i>Reciprocity and impartiality</i>	232
6. <i>Conclusion</i>	243

Chapter 6: Rawlsian Contractualism, Actual Contractualism, and Left-Libertarianism	
1. <i>The hybrid view: an overview</i>	246
2. <i>The hybrid view and the actual contractualist tradition</i>	253
3. <i>The hybrid view and left-libertarianism</i>	258
4. <i>Conclusion</i>	270
 Bibliography.....	272

Acknowledgments

Many people—philosophers, friends, and philosopher friends alike—have helped me over the years that I have worked on this thesis, and I am grateful to them all. I would, however, like to single out the following among them for special thanks.

My supervisors, Michael Otsuka and Véronique Munoz-Dardé, have been unfailingly patient, kind, and encouraging in manner as well as penetrating, constructive, and thorough in criticism and generous with their time. I have learned a huge amount from them and count myself unusually lucky to have been placed in their philosophical care. I am deeply grateful to both. I would also like to thank Jonathan Wolff and Mark Kalderon, under whose supervision I worked for a brief period, and each of whom contributed much to my understanding both of Rawls and of what I wanted to say about Rawls.

The UCL philosophy department as a whole has provided me with a highly supportive and intellectually stimulating environment in which to develop my thinking. I have the impression that the level of interest that the department takes in the philosophical and general well-being of its students is rare. I am grateful and proud to have spent my time as a postgraduate there.

I would in addition like to thank my long-suffering yet uncomplaining partner Ella Heeks. She has supported me intellectually, emotionally, financially, and, on occasion, physically for many years. I couldn't have done anything, much less produce a thesis, without her. I am especially grateful to her for her selflessness over the final few months of writing.

My parents, too, have been unstinting in their support. They encouraged my interest in philosophy to start with and have been interested, enthusiastic, generous

sponsors of it ever since. Moreover, they have been as wonderful a set of parents as any son could hope to have. I can't really thank them enough for that.

Adequate thanks for all the other people who have helped me would take up more room than I have here. I am, as I say, grateful to them all. Since none of them is likely to read this thesis, my hope that they won't be offended to be thanked only inadequately will probably be fulfilled.

The Arts and Humanities Research Council (formerly the Arts and Humanities Research Board) provided generous funding for my doctoral studies. I am grateful for its support.

This thesis is dedicated to the memory of my father, who died while I was writing it.

Introduction

1. Contractualism and the individualistic ideal of justification	8
2. Rawls and hybrid contractualism	14
3. Hybrid contractualism and egalitarianism	18

1. *Contractualism and the individualistic ideal of justification*

Contractualists about justice are moved by an ideal of individualistic justification.

They suppose that at least part of the story about why some conception of justice is the right one, the one that we ought to implement in our society, is that it satisfies the ideal. But there is more than one way to specify what it takes to satisfy it. This thesis considers two familiar but contrasting possible specifications and argues that there is a way to combine them into a single specification (the 'hybrid' view) which manifests the attractions of both. It suggests that this specification is to be found in, and offers us the right way to interpret, the later views of John Rawls, the most important political philosopher of the twentieth century.

The contractualist ideal in all its forms starts, unsurprisingly, from the idea of a contract. When two people make a contract in ordinary circumstances, something that one of them can say to the other if she fears that the other may be considering breaching the terms of the contract is: you *signed* the contract. Of course, she may also appeal to prudential considerations such as the fact that she can sue for breach of a lawful contract, or the fact that the terms of the contract are in both parties' interests. But the appeal to the person's having signed the contract contrasts with these considerations in that it ordinarily justifies the supposition that the terms of

the contract have normative force for the signatories. Prudential considerations may make it rational to act in accordance with those terms, but the fact of the individual's having actually signed the contract seems, in itself, also to justify a moral obligation on the part of the signatories. Now, the appeal to the fact of a person's signature on a contract is an appeal to the idea that *that person alone* incurred an obligation in signing. So a contract creates obligations to abide by its terms for each of and only its signatories, and in each signatory's case only because *she* signed it. In that sense, the normative force of a contract's terms is justified *individualistically*: it is justified, one by one, to each signatory by appeal to something which is specific to her. By contrast, we usually take such things as general principles of rationality (e.g., the principle that one should will the necessary means to one's end) to have normative force with respect to any given individual irrespective of any particular facts about her. There is no 'one by one' justification of the norms of rationality.

In the contractualist approaches to justice that I consider in Part I, this individualistic idea of justification is applied to the question how to justify (in particular) the terms which are coercively enforced by states. Those terms are normally taken to apply to all members of society, so contractualists seek their justification, one by one, with respect to all members of society: they seek, in a sense, as many justifications as there are members. The simplest application of the individualistic ideal to justice imagines that the coercively enforced terms just *are* the terms of a real contract: the social contract. Contractualists adopting this approach have standardly tried to show that these have been accepted by each person—either expressly, with a signature or declaration, or tacitly. More generally, they argue that the only just terms are those which are actually accepted by each person. This approach, which I call 'actual contractualism', is most famously associated with Hobbes and Locke. Its great attraction arises from the fact that it allows us to

justify the coercive enforcement of society's terms of cooperation in precisely the same way that a contract's terms are justified to its signatories: 'look,' we can say, '*you accept(ed) these terms*'. And it allows us make this appeal to every single person to whom the terms apply. Why is this so attractive? Well, first, it has the virtue of familiarity and simplicity. We use this form of justification in our everyday lives whenever we invoke a past agreement in order to justify an appeal to obligations that we take people to have. But it also has the attraction that it expresses the idea that each person should be sovereign with respect to the obligations that she has: she should only have those that she has in some sense taken on. If the terms of a social contract can be justified to each person by appeal to something that she *did*, an exercise of her agency, then imposing those terms upon her is consistent with seeing her as the creator of her moral world (at least in the normative areas in which those terms are so justified). This, I think, is what lies at the root of actual contractualism's attraction.

In Chapter 1 I analyse the fundamental presuppositions and the various forms of actual contractualism. I start by contrasting 'legitimacy' and 'regulative conception' approaches. The former make the justice of a coercive social system conditional upon its reflecting some morally relevant aspect of the particular individuals that are members of the society. The latter do not, requiring instead that the coercive norms of a society conform to principles which are the conclusions of an independent moral argument. But these two can be combined in various ways, as when a theorist proposes, as a necessary condition of justice, both that the coercive norms reflect aspects of society's particular individuals and that they conform to principles which are the conclusions of an independent moral argument.

Actual contractualism is a legitimacy approach. The morally relevant aspect of the particular individuals that a just coercive system reflects is, of course, the fact of

each individual's having acted in a way that justifies the imposition of that system's norms. I argue that essential to actual contractualism is some conception of the 'moral background'—the state of affairs that obtains prior to anyone's having accepted the terms of any contract—as well as an account of what kind of acceptance justifies the imposition of norms. Many theorists also tell us what was or would be justified given their accounts of these features. I note that they could legitimately appeal to the relevant individuals' moral views in doing this, although most are normally understood to appeal rather to self-interest.

In the light of this analysis, I analyse the views of Hobbes, Locke, and their contemporary descendant Nozick. I argue that Hobbes proposes a form of actual contractualism which subjects every single norm in a system of norms—at least part of which is taken to be coercively enforceable—to the actual contractualist test. I call such an approach 'thoroughgoing' and claim that it has an special attraction not well accounted for by the value of autonomy which appears to underlie Lockean views (including Nozick's). Both the Hobbesian and the Lockean approaches see individuals as sovereign in a particular area of their lives, but autonomy adequately motivates only an approach which keeps that area within certain boundaries. So if I am right that the more thoroughgoing approach is attractive, something more must be said.

The actual contractualist approach is not the only contractualist approach to justice. Another approach is equally inspired by the ideal of individualistic justification. This second approach, however, does not envisage the coercively enforceable terms which govern our societal relations as the terms of any actual contract. Instead, it asks: what would be the terms of a contract which no one could reject? I call this approach 'modal contractualism', and it is the subject of most of Chapter 2. Obviously the answer to the question that modal contractualists

ask is determined by what's meant in saying that someone *could* or *couldn't* reject some contract. In some sense, a person *could* reject any contract, regardless of its content. But the fact that someone could in *some* sense reject any contract doesn't seem relevant to the justice of imposing any contract's terms, so modal contractualists ask about a more specific sense which does seem relevant. Their question becomes 'what would be the terms of a contract which no one could *reasonably* reject?'

The reasonableness of a rejection of some contract is understood to be a matter of (normative) fact. If you decide not to sign a contract which represents a fair solution to some dispute that we have, and your reason for not doing so is that you judge your bargaining position to be sufficiently strong that you could negotiate a contract that is less fair but more favourable to you, then (granted this description) it is a matter of fact, we suppose, that your rejection of the fair contract is *unreasonable*. This points to the great attraction of the modal contractualist interpretation of the individualistic ideal of justification: it promises an account of just terms for the regulation of our societal relations which are *objectively* reasonable. For the terms of the modal contractualist contract do not track merely what all parties *can* agree upon (that might be influenced by unfair bargaining advantages or failure to think through the likely outcomes of the contract) or even what all parties *take* to be reasonable (with the best will in the world people can fail to give others' claims appropriate weight). They track what *really is* reasonable.

We might worry, however, that this focus on what is objectively reasonable, what it is unreasonable for people not to accept, pulls us away from the individualistic justificatory ideal that was supposed to be the basis of the modal contractualist approach. In what way, now, are the terms of a contract that the modal contractualist judges to be just (because no one could reasonably reject it)

justified one by one, in a way that considers the particular situation of each person, to those to whom they apply? Isn't the contract justified to me simply because it is justified by the objective reasonableness of its terms? The situation seems to be analogous to the justification of some rational deduction to a given person: no one could rationally reject the deduction, so *she* should accept it. But as I noted above, there isn't anything particularly individualistic about the justification of rational norms.

My discussion of modal contractualism in Chapter 2 concentrates on the question what, if anything, is distinctively individualistic about modal contractualist justification. After surveying various possibilities, I conclude that the distinguishing feature is the 'Individualist Restriction', which constrains us to argue for the reasonableness of the rejection of possible contractual terms by appeal solely to grounds which are accurately described as those held by single individuals alone. This can be said to retain the 'one by one' character of contractualist justification because it means that the question we must ask in determining whether terms could reasonably be rejected is: how do a given single individual's grounds for preferring terms other than the proposed terms compare in turn with each other individual's grounds for preferring terms other than the first individual's preferred terms? We do not add up everyone's grounds for and against each possible set of terms and then pick those with the highest positive aggregate weight in favour. Rather, we try to find the terms the strongest individual grounds against which are weaker than the strongest individual grounds against any alternatives.

Thus we arrive at two competing interpretations of the contractualist ideal of individualistic justification, each with its own attractions. In the remainder of Chapter 2 I consider the question whether the two might be combined into a single 'hybrid contractualism' which manifests the attractions of both. I suggest that this

could indeed be done by identifying as a strong, objectively reasonable ground for the rejection of a contract's terms the fact that an individual *actually* rejects it. But this also creates certain puzzles that the hybrid contractualist must solve if her view is to be viable. Foremost among them is the following. Modal contractualists suppose that we have a very powerful reason to find and comply with principles that no one could reasonably reject. This is a component of their view. This powerful reason might plausibly be supposed to bring us to reconsider any views that we have which would otherwise lead us to reject principles that no one *could* reasonably reject. But if principles that no one could reasonably reject are principles that no one actually *does* reject, as the hybrid view holds, then that powerful reason is a reason to reconsider any views we might have that would lead us actually to reject principles that no one else actually rejects. There is thus a pressure for consensus. Depending on just how powerful we suppose the powerful reason to be, that pressure might be a pressure for consensus on *anything*. Justice might therefore be settled upon merely through coordination rather than by anything like a balancing of the kinds of considerations that we would want to inform its content. This is a puzzle for hybrid contractualists.

2. Rawls and hybrid contractualism

In Part II, our focus shifts from contractualist frameworks and their interpretations of the individualistic justificatory ideal to the later work of John Rawls. The 'hypothetical contractualism' of Rawls's early work is standardly interpreted as a form of modal contractualism, and in his later work Rawls indeed explicitly aligns himself with the modal contractualist views of T.M. Scanlon. However, I argue that

within his modal contractualist framework Rawls identifies as a ground for the reasonable rejection of any of a range of liberal political principles the fact that a person does not, in a certain sense, actually accept those principles. Appealing to the idea of 'the burdens of judgment'—factors which account for and explain the reasonableness of disagreement about fundamental moral matters—Rawls claims that we are compelled to accept the justice of liberal political principles which safeguard the basic freedoms of conscience, thought, speech, and religion, among others. But there are a range of possible liberal political principles, and given the burdens of judgment we cannot say that any one set in this range is more reasonable than another. The justice of imposing a specific set, therefore, is dependent on their actually being accepted—in an 'overlapping consensus'—by each reasonable person (where to be reasonable is to accept the necessity of liberal principles in light of the burdens of judgment and to seek to impose only principles that no one could reasonably reject). Thus we find a seed of actual contractualism, albeit heavily constrained, at the core of the Rawlsian modal contractualist view. This makes Rawls a hybrid theorist.

In fact, as I argue in Chapter 3, Rawls shouldn't constrain the actual contractualism of his hybrid theory as he does. The burdens of judgment, I claim, justify a much more wide-ranging actual contractualist test, one which requires not only that the specific liberal egalitarian principles of 'justice as fairness'—Rawls's favoured contractual terms—actually be accepted by each person, but also that liberal principles in general be accepted by each person. For the burdens of judgment explain the reasonableness of disagreement not only on non-political moral matters and on the specific form that a liberal political set of principles should take, but on liberal principles themselves. If the burdens of judgment show us that it is unreasonable to impose the principles of some moral view on someone

who does not accept it, then they show us that it is unreasonable to impose a liberal system on someone who does not accept it. The actual contractualism that Rawls takes the burdens of judgment to imply is wider-ranging than he realises.

In the remainder of Chapter 3 and in Chapter 4 I develop and defend the revised Rawlsian view that this leaves us with. I start by arguing that we can solve the puzzles that I raised for hybrid views in general at the end of Chapter 2. With regard to the puzzle I described a moment ago, I argue that a plausible construal of the import of the modal contractualist reason that we have to find and comply with principles that no one could reasonably reject is as follows: appropriate respect for that reason entails not that one is willing to jettison any views that one holds which might conflict with what no one else actually rejects, but that one is willing to reconsider carefully one's grounds for affirming them and perhaps alter them in the light of that. This creates a pressure towards consensus, but not towards a consensus that can be secured merely by settling on contract terms chosen at random. Individuals' moral starting points will make a difference. This solution to the puzzle also helps, as it turns out, to solve other puzzles which the hybrid view raises. It also lends the hybrid theory certain attractions that modal contractualists have claimed for modal contractualism without being able to show clearly why. Or so I maintain.

I also explain, at the end of Chapter 3, how what Rawls takes to be necessary for the justification of justice as fairness—that each person shares the fundamental elements out of which justice as fairness is constructed—is an acceptable construal of 'actual acceptance' from an actual contractualist point of view. My explanation appeals, first, to the idea that some form of *consent* is not an essential part of the distinctive actual contractualist form of justification (an idea I introduce in Chapter 1), and, second, to the idea that it is plausible to appeal to the exercise of agency

involved in affirming a moral outlook as a justification for the imposition of norms derived by elaborating elements of that outlook.

Three Rawlsian innovations are particularly pertinent for the defence of the hybrid view as I develop it. One of them is the ‘burdens of judgment’ basis for its actual contractualist component, which I have just discussed. The second is the emphasis that Rawls places on the *moral* character of individuals’ reasons for rejecting and accepting proposed contract terms. This is an idea which can also be found in traditional actual contractualist views (as I point out in Chapter 1) but which tends to be underemphasised. It is plausible to construe actual acceptance in an actual contractualist theory—especially a thoroughgoing actual contractualist theory—in terms of congruence not with individuals’ rational self-interest but with their considered moral views. Clearly this fits well with the burdens-of-judgment basis for actual contractualism. And it does not, I claim, undermine the fundamental attractions of actual contractualism in the first place.

The third Rawlsian innovation is what I call the ‘head-first’ approach to the actual contractualist legitimising process which is part of actual and hybrid contractualist views alike. The head-first approach asks whether the terms of some proposed contract would be actually accepted by those who are brought up under it, rather than whether they would be actually accepted by individuals in other circumstances such as the traditional actual contractualists’ ‘state of nature’. Although the head-first approach invites charges of indoctrination, which I spend the second half of Chapter 4 meeting, it is helpful in that it reduces our dependence on historical speculation as compared to traditional actual contractualists. Nevertheless the argument for any particular set of terms must, in a hybrid framework, depend on a great deal of empirical understanding and moral psychology. In the first half of Chapter 4 I bring together the various reasons that

Rawls offers as part of his ‘political constructivist’ strategy to show that the terms of justice as fairness would be able achieve the ideal of individualistic justification as it is interpreted by the hybrid framework.

This is not an easily achieved goal because justice as fairness, a liberal conception of justice, would engender a pluralistic society—one in which different individuals affirm a range of incompatible moral views. This partly explains why the Rawlsian view escapes charges of indoctrination: but it also makes it difficult to show that each individual brought up under the institutions of justice as fairness would come to endorse the terms of the social contract that it sets. The political constructivist strategy, as I argue, explains why various obstacles that we might expect to stand in the way of unanimous actual acceptance of any set of terms in such a pluralistic society do not in fact do so. But it remains to be shown all the same that an overlapping consensus would arise on justice as fairness.

3. Hybrid contractualism and egalitarianism

Justice as fairness articulates a strongly egalitarian social contract, in the sense that it prioritises equality in the distribution of material means, and this fact about it represents a greater barrier to the plausibility of supposing that it could pass the actual contractualist test that Rawls sets for it than does the fact that it also safeguards the basic liberties. This is for two reasons. First, traditional actual contractualist approaches to justice tend to have a built-in bias in favour of contracts safeguarding the basic liberties because it is part of their conception of the pre-societal state of nature which forms the backdrop to the contract that we are more or less equally free to do, think, and be as we wish. No one, it is supposed,

would be prepared to give up that freedom except insofar as doing so would eliminate conflict, and no one person is powerful enough to force a contract which is *inegalitarian* in its treatment of individual liberty. Since such conflict would be eliminated by restricting and enforcing each person's freedom so that it is fully compatible with each other person's equal freedom, it seems unlikely that any further restriction would be agreed to. Hence a bias in favour of the basic liberties. Now, given this bias in traditional actual contractualism, it seems more likely that Rawls's actual contractualism will have difficulty with its strong egalitarianism—something which it does *not* share with traditional actual contractualist views—than with its endorsement of the basic liberties.

The second reason to suppose that the strong egalitarianism will pose the greater barrier is as follows. I argue in Chapter 4 that part of Rawls's defence of the head-first approach against charges of indoctrination appeals to the fact that liberalism—i.e. political morality which endorses the basic liberties—represents, as a matter of historical fact, a point of convergence for societies which prior to becoming liberal societies were each organised around illiberal or non-liberal moral views which differed from society to society. Granted this fact, it seems empirically more plausible to posit a consensus on the basic liberties than it does to posit a consensus on strong egalitarianism. For strong egalitarianism is *not* something that we can so plausibly suppose, as a mere matter of historical fact, to be a point of convergence for even a range of liberal societies. Even if ultimately it will turn out to be so, we have not reached that point yet.

So in Chapter 5 I concentrate on the case for an overlapping consensus on justice as fairness's egalitarianism rather than the case for its liberalism. I offer exegesis of Rawls's well-known 'informal argument' for the egalitarian difference principle which stresses its roots in a conception of fairness which can plausibly be

attributed to members of the kind of liberal democratic society that justice as fairness would engender. The argument starts with the arbitrariness of the fact that each of us has abilities and endowments in different degrees. It then appeals to the idea that given the necessity of the involvement (not always in straightforwardly productive ways) of each of us in the particular scheme of cooperation in which we are engaged to the production of the specific set of goods which that particular scheme produces, and given also the arbitrariness of the fact that *this* is the scheme of cooperation which produces the goods that we want, each person has a *prima facie* claim to an equal share. Finally, this equal share is finessed into the share mandated by the difference principle as a reflection of the fact that each of us could do better with respect to the baseline of an equal share if we permit some inequalities.

This is not a novel reading. But it is worth offering in order to stress the plausibility of the claim that the intuitions of fairness upon which it is based are common and uncontroversial. And, perhaps more importantly, it is worth offering in order to dispel the idea that Rawls draws upon an idea of fairness as impartiality which is much less plausibly attributed to each person in liberal societies. I try to bring out at the end of the chapter exactly why motives of impartiality with their cosmopolitan implications are not required to support justice as fairness's egalitarianism despite objections which arise in the light of the possibility that individuals might 'shop around' for alternative cooperative schemes in which they do better. This explanation highlights once again the importance of the moral character of individuals' reasons for actual acceptance in the Rawlsian hybrid framework.

In the final chapter I summarise the hybrid contractualism that I have developed and defended before considering its similarities and contrasts with the actual contractualist tradition. (I concentrate on the actual contractualist tradition

because the modal interpretation of Rawls's contractualism is well-known and, in any case, there is no great tradition of modal contractualism with which to compare the hybrid view.) I note the unusual (for a philosophical political theory) empirical emphasis of the hybrid view but suggest that this should be thought of as in keeping with the actual contractualist tradition. I also note that the hybrid interpretation of Rawls offers an interesting and plausible understanding of his view that political philosophy should be 'realistically utopian'.

I end with a brief comparison of the hybrid view with another recent attempt to reconcile the attractions of actual contractualism and egalitarianism: left-libertarianism. Left-libertarians endorse a right of self-ownership which, among other things, implies an actual contractualist right against the imposition of the terms of contracts which they do not actually accept. But they also endorse a theory of world-ownership which places strongly egalitarian conditions on anyone's coming to own any tract of the world outside her own person. I argue that while left-libertarianism represents an improvement on right-libertarianism because of its egalitarian rather than *laissez-faire* treatment of world-ownership, it represents at best no improvement at all in terms of the reach of its actual contractualist test, which is no more thoroughgoing than that of right-libertarians. Neither left- nor right-libertarians subject their favoured original principles of world-ownership to individualistic justification. Left-libertarianism may even represent a step backwards from the point of view of actual contractualism if, unlike right-libertarianism, it does not permit the imposition of alternatives to its egalitarian original principles of world-ownership even if each person actually accepts them. The hybrid view, by contrast, represents an improvement on libertarianism both in terms of its egalitarianism and in terms of the thoroughgoingness of its actual contractualist test. For various reasons which I note, it may seem to be a rather unsatisfying theory all

the same. I hold, however, that this is a reflection of the deep difficulties involved in satisfying both the contractualist ideal and the egalitarian intuitions that many mainstream political philosophers endorse.

PART I:
CONTRACTUALIST
FRAMEWORKS

Chapter 1: Actual Contractualism

1. Actual contractualism	24
2. Actual contractualism and approaches to justice	26
3. Two variables: scope and acceptance	35
4. The shape of actual contractualism	39
5. Hobbes, Locke, Nozick, and actual contractualism	48
6. Autonomy and the appeal of actual contractualism	62

1. *Actual contractualism*

One way to interpret the fundamental contractualist ideal of individualistic justification is as follows. A moral conception satisfies the ideal, we can say, if and only if each relevant individual actually accepts it. I shall call this the ‘actual contractualist’ interpretation of the contractualist ideal. Actual contractualist theories of justice, then, make a moral conception’s satisfaction of the ideal in its actual contractualist interpretation necessary for the realisation of justice.

Eventually I shall defend an actual contractualist interpretation of John Rawls’s contractualism. I shall begin by sketching in more detail the general features of actual contractualist theories of justice.

The description I just gave of the actual contractualist interpretation of the ideal of individualistic justification involved three main ideas. First, the idea of a moral conception, the kind of thing that satisfies the ideal. Second, the idea of a relevant individual, whose actual acceptance of the moral conception in question is necessary and sufficient for the satisfaction of the ideal. And third, the idea of actual

acceptance itself. The idea of a moral conception is simply the idea of a coherent set of normative principles and conceptions of persons and institutions and their roles and aims, together with an account of their basis.¹ In a moral conception appropriate for a theory of (societal, rather than intersocietal) justice the principles and conceptions of persons and institutions and so on are those invoked in order to spell out requirements of justice. Which of the requirements that we face are requirements of justice and which are not is an awkward question which I skate over here. I shall say that any correct principle of coercive enforcement of some other norm makes that other norm's requirements *ipso facto* requirements of justice, and that it is also sufficient for requirements to be requirements of justice that they are created by norms which are part of a moral conception that endorses the coercive enforcement of at least *some* norms even if not those in question.² Any moral conception which sets out institutions and principles which create

¹ The idea of a moral conception I take from Rawls, although he never specifies precisely what he understands by it. The idea is nonetheless fairly clear from his discussion of 'political conceptions', which constitute a subset of all moral conceptions. See *PL*, pp. 11-15. Rawls's Kantian constructivism offers another example: see "Kantian Constructivism in Moral Theory" in Rawls's *Collected Papers*, p. 305.

² Although affirmation of any set of norms usually implies the appropriateness of the employment of some kind of pressure to conform, coercion is a particularly serious kind of pressure to conform which makes the cost of non-conformity higher than do other kinds. This goes some way towards justifying the focus on coercion as the key feature of requirements of justice. Cf. Samuel Scheffler, "Is the Basic Structure Basic?", p. 124: "although I am not sure that Rawls would or should construe the basic structure in purely coercive terms, there is one obvious reason for doing so that does not seem arbitrary, namely, that the coercive structure is *coercive*." It seems obvious to Scheffler that coercion is important in a way that moral disapproval, say, is not, and that this would be a reason for concentrating primarily on coercive institutions.

requirements of justice is a moral conception appropriate for a theory of justice.³

From now on, when I use the phrase ‘moral conception’—and I shall use it often—

I shall mean ‘a moral conception appropriate for a theory of justice’.

I shall look at the idea of the relevant individual and the idea of actual acceptance implicit in the actual contractualist interpretation of the fundamental contractualist ideal in a moment. First, a word about the connections between actual contractualism, justice, and legitimacy. As I shall frame the issue, actual contractualist theories of justice can be either what I call ‘pure legitimacy theories’ or ‘combined theories’ but not ‘regulative conception’ theories. In the next section I explain this way of framing things in detail. Then, after I have said more about relevant individuals and actual acceptance, I describe the general structure of actual contractualism.

2. *Actual contractualism and approaches to justice*⁴

We ought to act in accordance with those institutions and principles of justice in our society which are justified. If a principle which forbids harming in most circumstances is justified, for example, we ought not to harm in those circumstances. Some institutions and principles may be coercively enforced, and

³ Whether these institutions are those of Robert Nozick’s minimal state, those of Rawls’s ‘basic structure’, or otherwise will depend on the content of the moral conception in question. This is not predetermined by the conception of which requirements are requirements of justice that I propose here.

⁴ My account in this section owes a great deal to that of A. John Simmons in his helpful discussion in ‘*Justification and legitimacy*’.

the coercive enforcement of them must be part of what is justified. Coercive enforcement is not an extra which comes automatically when the subject matter of principles is within a certain range. It is optional, and if it is included it too must be justified.

The way in which theories of justice justify the principles, institutions, and coercive enforcement of institutions and principles varies from theory to theory. Suppose that a tyrannical dictatorship, unjust in anyone's book, is transformed instantaneously so that its institutions and principles, including its laws, are those advocated by some theory of justice. Some theories of justice would say of this miraculously transformed society that we could know from this alone that it had become, instantaneously, a just one. They would say this regardless of the relations or attitudes or behaviour of the members of society beyond the disposition to act in accordance with the institutions and principles. Call theories of justice that would say this *pure regulative conception* theories of justice. One example of a pure regulative conception theory is utilitarianism. (Non-rule) utilitarians think that if the (fully conformed with) institutions and principles of a society are such that they produce the greatest happiness of the greatest number, then that society is just. If the miraculous transformation involved the sudden establishment of such institutions and principles, it would in that instant become, on a utilitarian theory, just.

Now, we suppose, I think, that at least *part* of the reason that we have to act in accordance with just institutions and principles is that these institutions and principles are those of *our* society. But pure regulative conception theories of justice don't make this fact a central component of their explanations of why we ought to act in accordance with the institutions and principles that they advocate. They tell us that these institutions and principles are the right ones, simply, and that therefore we ought to be implementing and acting in accordance with them in our society.

If we thought that all just individual action were entirely determinable in advance of the establishment of any institutions and principles to prescribe it, then it wouldn't make much sense to suppose that even part of the reason that we have to act in accordance with just institutions and principles is that they are those of our society. If we could know what justice would require of Jack in every conceivable situation that Jack could find himself in, and without consulting the institutions and principles established in any society that Jack happened to be a member of, then what role could the fact that the institutions and principles of the society of which Jack is a member *are* those of that society possibly play in accounting for the reason that Jack has to act in accordance with them? Since we do suppose that part of the reason that we have to act in accordance with just institutions and principles is that they are those of our society, it can't be the case that we think that just individual action is *entirely* determinable in advance of the establishment of any institutions and principles to prescribe it.

But this on its own doesn't tell against pure regulative conception theories, for among the institutions they prescribe might be legislative institutions with the discretion to legislate a range of things within certain constraints. A just society would have to have these institutions, but we may know this without knowing precisely what will be required of individuals in a just society, because we may not know which of a range of possible rules that are consistent with justice have been prescribed by the institutions. So, for example, a theory of justice might (albeit implausibly) say that whether or not there should be an absolute prohibition on striking others with one's fists should depend upon how long it takes the tallest member of society to count everyone else. If it takes her longer than a certain amount of time, then there should be an absolute prohibition. If not, there should not. In this case, just individual action isn't entirely determinable in advance of the

establishment of the institution which involves the tallest member of society's counting everyone else. So pure regulative conceptions aren't inconsistent with the supposition that part of the reason that we have to act in accordance with just institutions and principles is that they are those of our society. They are therefore not ruled out on that basis as plausible conceptions of justice.

So if we still think that some part of the explanation why we ought to act in accordance with just institutions and principles goes missing in a pure regulative conception theory we had better refine our understanding of what it is. I propose that the point is not merely that the relevant institutions and principles are those of *our* society but that they reflect, in some *important* sense, the particular, distinct persons we are. That is why the fact that the pure regulative conception theory that I just imagined can't prescribe just action without reference to the make-up of the relevant society isn't enough for it to respond adequately to the intuition. The time it takes the tallest person in society to count everyone else is not reflective of us in any *important* sense.

Of the theories of justice which are not pure regulative conception theories, those whose explanations of why we ought to act in accordance with just institutions and principles make reference *solely* to facts about the particular persons we are in the relevant sense I shall call *pure legitimacy* theories of justice. These theories take these facts to be relevant because they are said to give rise by themselves to obligations to comply with the institutions and principles of society. They make those institutions and principles, including principles licensing coercive enforcement, just. Your society is just if and only if its members have the obligations generated in the relevant way and regardless of whether the principles and institutions of its moral conception meet any independently specified conditions. In principle, any set of institutions and principles could be just on a

pure legitimacy theory, although in each case this will of course depend on exactly how obligations of justice are supposed to depend upon facts about the relevant persons. Theories of justice which make it a necessary and sufficient condition for the justice of a society that it be ruled (using force) by a person with hereditary title to the throne are obviously pure legitimacy theories. That person's subjects ought to comply with the legislation of their monarch simply because they are subjects of that monarch (a fact about them whether they are ruled by her or not). Their subjecthood is a special obligation-generating fact about them. If they were subjects of some other person, then *this* monarch's rule would not be just, even if what she legislated was exactly what the rightful monarch would legislate. (Compare the obligation that I might be taken to have to obey my parents.)

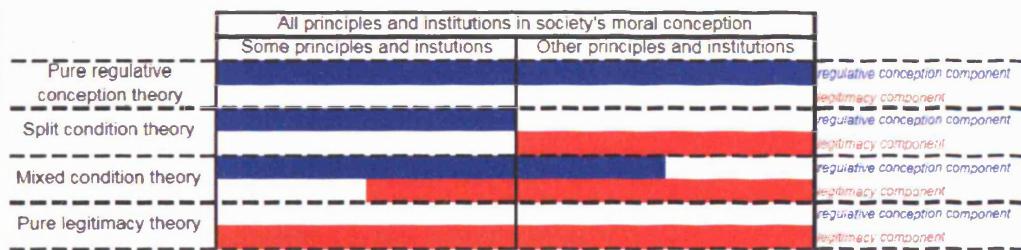
Pure regulative conception theories and pure legitimacy theories lie at opposite ends of a spectrum. In between are theories of justice which combine the two (see Figures 1 and 2 below). This can be done in three different ways. The first and perhaps most obvious way to combine them is to subject some of the institutions and principles of a moral conception to a pure legitimacy test and others to a pure regulative conception test. So, for example, we might say that this society's constitution ought, if it is to be just, to meet a set of conditions determined independently of any facts about these particular persons who compose this society. But at the same time we might say that the laws not pertaining to the constitution ought to be determined by majority vote, conceived as a means of making known the general will. Or we might say that the person with legitimate claim to the throne ought to govern within a clearly bounded area outside which the content of the laws should be independent of her legislation. I shall call theories which combine regulative conception and legitimacy views in this way *split condition theories*.

The second way to combine them is to offer an account of why we ought to behave in accordance with just institutions and principles which sees us as obliged only because of *both* facts about our persons *and* the fact that those institutions and principles meet conditions that those of a moral conception ought to meet independently of facts about us. Only if a society's moral conception meets the relevant conditions *and* each of its members has obligations generated in the way that the account specifies can the society be said to be just. So, for example, we might say that every institution and principle in a just society ought to reflect the general will *and* ought to be justified according to some criterion independent of facts about the particular, distinct persons who compose society. If the moral conception reflects the general will but does not meet the independent criterion it is not just. If the moral conception meets the independent criterion but does not reflect the general will it is not just. I shall call theories which combine regulative conception and legitimacy views in this way *dual condition theories*.

The third way to combine the two views is a mixture of the first two. According to *mixed condition theories*, we ought to behave in accordance with *some* just institutions and principles in society's moral conception because they meet certain conditions *à la* pure regulative conception theories. Or we ought to behave in accordance with them because they are appropriately reflective of the distinct persons we are *à la* pure legitimacy theories. Or we ought to behave in accordance with some because they meet certain conditions and with others because they reflect the distinct persons that we are. In addition, we ought to behave in accordance with *still others* *both* because they meet certain conditions *and* because they are appropriately reflective of the particular, distinct persons we are. So, for example, someone might advocate a theory which held that a certain area of legislation, outside which the laws should be set in stone, should be governed by a person with

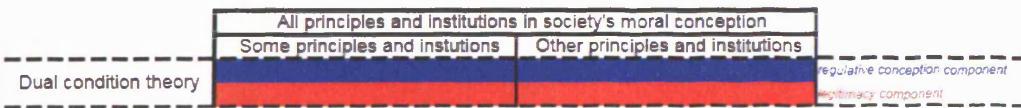
legitimate claim to the throne but so that her decrees were coextensive with what is justified independently of her will. This would be a mixed condition theory.

Figure 1.



Pure regulative conception theories subject all elements of a society's moral conception to a regulative conception test (blue) only. Pure legitimacy theories subject all elements of a society's moral conception to a legitimacy test (red) only. Split condition theories and mixed condition theories subject some elements of a society's moral conception to one test, some to the other. Mixed condition theories also subject some elements to both tests.

Figure 2.



Dual condition theories subject all elements of a society's moral conception to both a regulative conception test (blue) and a legitimacy test (red).

Now, actual contractualist theories may be pure legitimacy theories or they may be combined theories (i.e. either split, dual, or mixed condition theories). But they

cannot be pure regulative conception theories, since they always include the requirement that a just society satisfy the ideal of individualistic justification—in its actual contractualist interpretation—of the relevant moral conception. As we saw at the outset, the actual contractualist interpretation of the ideal involves relevant individuals' actual acceptance of the moral conception. So justice is not attainable without actual acceptance, and the explanation of why a moral conception is just makes ineliminable reference to a fact about the particular, distinct persons that compose society: the fact of their actual acceptance. This is why actual contractualists cannot be pure regulative conception theorists.

Actual contractualist theories which are also pure legitimacy theories set no further conditions for the justice of any given society beyond that its moral conception should meet the actual contractualist interpretation of the contractualist ideal. Actual contractualist theories which are combined theories set independent conditions on at least some of the institutions and principles of society's moral conception and require the satisfaction of the contractualist ideal with respect to at least some others: they have a regulative conception component. The most demanding form of actual contractualism is dual condition actual contractualism. Whereas theories with a *pure* regulative conception component concentrate only on the content of at least some institutions and principles, and theories with a *pure* legitimacy component concentrate only on obligation-generating facts about members of society with respect to at least some institutions and principles, dual condition actual contractualists set dual conditions on the entire content of moral conceptions. These conditions are: that they fit with some non-legitimacy based justification (the regulative conception component); and that they meet the actual contractualist interpretation of the contractualist ideal (the legitimacy component). In setting themselves this challenge, dual condition actual contractualists

acknowledge the appeal of both pure legitimacy theories and pure regulative conception theories. On the one hand, they respond, as pure regulative conception theories do, to our conviction that justice with respect to *any* institution or principle is a matter at least to some extent independent of what anyone thinks is just at a given time. All of the people, as we might put it, are mistaken some of the time; even those who lose out under some moral conception may collude in their own oppression. Indoctrination is possible and powerful, so we must look beyond the views and actions of those subject to given institutions and principles in order to determine whether the society which implements the relevant moral conception is just. To grant that any institution or principle meeting a legitimacy criterion (at least on some versions) is just is to ignore this conviction.

On the other hand, dual condition actual contractualists acknowledge more fundamentally the supposition that part of the reason that we have to act in accordance with just institutions and principles is that they reflect the distinct, particular persons that we are. They also have, moreover, the advantage of an *attractive* account of the ground of the obligations to comply with just institutions and principles that we take ourselves to have. A dual condition monarchist theory, for example, says that a society is just if and only if the person who is entitled to the throne implements institutions and principles which themselves meet certain further conditions. (Obviously the regulative conception component must be consistent with monarchy.) But the necessity of the rule of a monarch who has a legitimate claim to the throne is unconvincing as a component in a theory of justice. Universal actual acceptance of a moral conception is much more appealing as the basis of legitimacy than the ancestry of the person who imposes the rules, whatever they are.

A dual condition theory whose account of legitimacy is actual contractualist looks to be an appealing framework, then. The Rawlsian view that I am going to

defend operates within it. For the most part of this thesis, however, and for the entirety of this chapter, I shall be concentrating on the actual contractualist component of this kind of theory. This is because the regulative conception component—the argument for the principles of justice that proceeds from the original position—of Rawls's theory of justice is well-known. (I do discuss it, nevertheless, in more detail in Chapter 5.) I consider the question of the basis of actual contractualism's appeal further on in this chapter. But first I want to clarify the ideas of actual acceptance and of the relevant individual implicit in the actual contractualist interpretation of the ideal of individualistic justification.

3. *Two variables: scope and acceptance*

The actual contractualist interpretation of the contractualist ideal of individualistic justification, recall, says that a moral conception satisfies the ideal if and only if each relevant individual actually accepts it. The content of actual contractualist theories changes according to two variables representing a range of ways to think about the ideas of actual acceptance and of the relevant individuals whose actual acceptance matters which we find in this interpretation of the ideal. The first variable, which I'll simply call 'acceptance', concerns the construal of the idea of actual acceptance of a moral conception. When has a conception *actually* been accepted by someone? This is a question about how we interpret empirical facts. We might say any of a range of things. At one extreme, we might say that a moral conception has actually been accepted by someone if she is not making violent efforts to implement some alternative. Actual contractualist justification of all sorts of conceptions may be easy to secure if we say this. For example, it may be secured in a society in which a few

live like kings and the rest are too busy scraping together a living to try to overthrow the current regime. At the other extreme, we might say that a moral conception has actually been accepted by someone only if she freely puts her signature to a declaration of consent to it after years of studying the arguments for it and for the alternatives to it and she repeats such signed declaration of consent on a daily basis. Actual contractualist justification of a moral conception of any kind is extremely unlikely to be feasible if we say this. So in determining the content of actual contractualism it is necessary to decide upon and argue for one of many ways to understand acceptance.

What is fundamental to all construals of the variable of acceptance, however, is the idea that individuals should somehow have control over the imposition on them of norms in the areas we take to be subject to the actual contractualist test. The individual is treated as an agent whose *actual* actions alone, construed as acceptance, should determine whether and which norms constrain her in those areas. This is the source of actual contractualists' ability to justify the terms of their favoured contract by saying to each individual: '*you* accepted it'—the basic idea in the actual contractualist interpretation of the individualistic ideal of justification. However we construe the acceptance, for this to have any force at all as a distinctive form of justification it must state, or at least infallibly track, the fact of the relevant individual's actual exercise of her agency. We are, that is, looking for something that can genuinely be understood as an exercise of agency rather than, say, the outcome of a compulsion or the only metaphysical possibility. Otherwise to say '*you* accepted it' doesn't have force distinct from 'that's who you are', the justificatory basis of legitimacy views in general, or from 'that's the way things (morally) are', the justificatory basis, roughly speaking, of regulative conception views.

But notice that the form of actual contractualist justification as I've just specified it does not imply that acceptance must be construed so that it involves individuals' consciously accepting the terms of the contract in an everyday sense (i.e. by consenting to or agreeing to them). What marks out actual contractualist justification, I suggest, is that it makes the legitimacy of the imposition of norms upon an individual dependent upon an exercise of that individual's agency which can then be appealed to in justifying the contract to her. Structurally speaking, this is sufficient to distinguish it from legitimacy approaches more widely and of course also regulative conception approaches. I see no reason to go further and insist that the exercise of individual agency that it appeals to be *consent* in some form.⁵ That does not, however, mean that any exercise of agency will do as a construal of acceptance: it makes no sense at all, for example, to suppose that your whistling on your way to work tomorrow constitutes acceptance of my coercion of you in accordance with a moral conception that I pick out at random. *Plausible* actual contractualist accounts must *plausibly* connect their construals of acceptance to the moral conception accepted and to their concern with actual contractualist justification in the first place. Express consent as a construal of actual acceptance seems plausible because it is part of our ordinary practice of taking on new obligations. Tacit consent is also part of that practice and so seems plausible too.⁶

⁵ Cf. Rawls, *Collected Papers*, p. 210, n. 16, where Rawls suggests that Locke (undoubtedly an actual contractualist) uses tacit consent in a sense which means that for him "an obligation to obey the rules laid down by a legitimate political authority can arise independently of a performative act, whether express or tacit, although not without a prior voluntary act."

⁶ Both of these are readily understood as exercises of individual agency. In the case of tacit consent, surveys of conditions necessary for someone's behaviour to be construed as tacit consent make it clear that the possibility of *dissent* must have been available and understood—which implies that the

But all the same we should not suppose that traditional construals of acceptance such as free, informed consent are the only plausible construals that could be consistent with actual contractualism. Other actual exercises of individual agency might also be appealed to to justify the imposition of norms; doubts then will be about their plausibility rather than their consistency with actual contractualism.

Further on I shall defend a non-consent based construal.

The second variable, scope, concerns the constituency to which justification is taken to be owing. This may seem surprising. Shouldn't constituency be not a variable but a constant, viz., all individuals who will be obliged to comply with the institutions and principles of moral conception in question? But if the constituency is delimited thus, actual contractualism may fail to live up to the contractualist ideal almost regardless of the value it assigns to the first variable. Some individuals may not be capable of giving actual acceptance on any assignment of the variable of acceptance which even remotely tracks whatever it is that makes us interested in it in the first place. Yet we may still want to say that these individuals are subject to the authority implementing the moral conception in question. We might think, for example, that people who are insane or people in comas or people who are pathologically incapable of complying with norms are obliged to comply with the institutions and principles of the right moral conception. But such individuals may not be capable of accepting those principles on any understanding of actual acceptance that we find it plausible to apply in general.

Why isn't this problem of scope simply a counterexample to actual contractualism? Actual contractualist theories are supposed to offer an articulation

failure to dissent counts as an exercise of the agent's choice. See e.g. Simmons, *Moral Principles and Political Obligations*, pp. 80-3.

of the ideal of individualistic justification. On any plausible formulation they may fail to secure the actual acceptance of each of those who will be subject to rule according to the principles of the moral conception in question. So why shouldn't we reject actual contractualism? The answer is that actual contractualism—and only actual contractualism—responds, even in a form which restricts the constituency to whom justification is taken to be owing, to concerns which its proponents take to be a reason to endorse the ideal of individualistic justification at all. The extent of the restriction, moreover, need not be determined in an *ad hoc* manner but as a result of reasoning which connects with this fundamental concern. I shall say more about this below.

4. *The shape of actual contractualism*

With many of their components clarified in the preceding sections, I shall now try to give a brief sketch of the general structure of actual contractualist theories. As we have seen, such theories say that a moral conception or some part of it must meet the condition of actual acceptance by each relevant individual. This is their defining characteristic, and their central structural feature is the transition via actual acceptance as at least a necessary condition of the existence of our obligations to comply with that moral conception's institutions and principles. Any plausible such theory must offer understandings of the meaning of 'actual acceptance' and 'relevant individual'—assignments of what I have called the variables of acceptance and scope—which are consonant with the underlying attraction of actual contractualism's interpretation of the ideal of individualistic justification. It must be made clear what is involved in actually accepting a moral conception.

Now notice that in postulating the necessary condition of actual acceptance, actual contractualists take each relevant individual to be, in effect, the bearer of a moral right against, at the very least, coercion to behave in accordance with some institutions and principles which she does not actually accept. (This is minimally a right against *coercion* because as I stipulated in section 1 above, a moral conception spells out at least coercively enforceable institutions and principles.) This right forms at least part of the *moral backdrop* for the actual contractualist theory. A correlate of the demand for actual acceptance, this moral backdrop is the second of actual contractualism's main structural features. The transition to an actually accepted moral conception must take place against this moral backdrop.

Such theories need not say anything more, and some pure legitimacy versions might not. They might of course describe the (relevant part of the) moral conception that they think most likely to achieve acceptance and, therefore, to be that of a just society, but this description would do no normative work. If no one actually does accept such a conception, the fact that it seems in advance more likely to achieve legitimacy through actual acceptance than something that everyone actually does accept does not make it right to impose it.

But the moral backdrop needn't be restricted to a right against coercion to behave in accordance with principles that one doesn't accept. An important way in which it might be expanded beyond such a restriction is this: the moral backdrop itself is taken to include certain institutions and principles of justice.⁷ This is not consistent, however, with a pure legitimacy theory. The transition via actual

⁷ It is reasonable to talk of justice with regard to a society subject to the original moral conception if the moral conception includes principles mandating the coercive enforcement of institutions and principles. The lack of a state—understood as having a monopoly on the legitimate use of force—does not in my view render inapplicable talk of justice. See Section 1 above.

contractualist acceptance to the accepted moral conception takes place against the background of another moral conception some of which, at least, is not itself subjected to the actual contractualist test. How this works out depends on the kinds of restrictions that the principles of the background moral conception are taken to impose. Suppose that the institutions and principles that the background moral conception involves are taken to impose absolute restrictions on what can be accepted, so that society cannot be just except if these institutions and principles are part of its moral conception. In that case society is just prior to the actual contractualist transition to a new moral conception and it may not be afterwards. If the newly accepted moral conception violates the original conception's restrictions, it's not just. If it doesn't, it is just, but to the extent that the new moral conception simply repeats the restrictions of the original one, actual acceptance is doing no work at all. The original moral conception may, however, leave room for discretion over various domains of conduct—it may say nothing about them except that regulation of them must meet the actual contractualist test—in much the same way that an unalterable constitution requiring democratic decision-making leaves room for collective discretion over those domains about which it permits democratic decisions to be made. In that case in those domains the actual contractualist requirement does do work in justifying institutions and principles governing those domains. The picture as a whole is (for all I have said) a split theory: some aspects of the advocated moral conception subjected to a pure regulative conception test, others to a pure legitimacy test. That both tests are met suffices for the justice of society.

So the original moral conception, which is not subject to the actual contractualist test, determines the content of the new moral conception directly, because it simply specifies part of the final moral conception. And (depending on

the grounds that inform individuals' actual acceptance) it may do so indirectly, because it may exercise an influence on the content of the other part. So, for example, the original moral conception might stipulate robust rights of self-ownership⁸ but say nothing about acquisition or ownership of property except that the relevant institutions and principles must meet the actual contractualist test. But the rights of self-ownership may inform the content of what each person actually accepts with regard to institutions and principles of property acquisition and ownership. For instance, the physically stronger may not intimidate the weaker into accepting unequal rights of acquisition. To the extent that without such intimidation the rights of acquisition will be more equal, the original moral conception informs the content of the final moral conception which includes institutions and principles of property acquisition.

These are the possibilities that present themselves if the moral backdrop in an actual contractualist theory involves a moral conception whose restrictions are taken to be absolute. Suppose, now, that the institutions and principles of the background moral conception are taken to impose restrictions which are not (all) absolute, so that the moral conception legitimised by the actual contractualist transition may justly involve restrictions which contravene norms imposed under the original conception. In this case, a society is just only if to the extent that its moral conception deviates from the original background moral conception, it is actually accepted by all members of society against the backdrop of that moral conception. The endorsement of the original moral conception as part of the actual contractualist's moral backdrop has the effect of specifying not only a set of institutions and principles which are just even if they are not accepted by each

⁸ See Michael Otsuka, *Libertarianism without Inequality*, p. 32.

person but also, as with the absolutely restricting background moral conception view, a set of bargaining positions which may inform, in the way I described a moment ago, the content of any just deviations from the content of the original moral conception through actual acceptance. For, as I claimed above, the moral backdrop must be in place as part of the actual contractualist requirement of actual acceptance of the moral conception that is ultimately produced and endorsed. What people accept against this background, rather than against any other, is, then, what is relevant.

In either case the expanded backdrop may also inform the content of the new moral conception in another way that it is worth bringing out. As we have just seen, since the actual contractualist transition *must* take place against the just background of the original moral conception, individuals' bargaining positions are set by the content of the original moral conception. Now how things are for individuals against the moral background determines their bargaining positions regardless of whether it is expanded or not. However, when it *is* expanded there is a certain moral endorsement on the part of the theorist of those bargaining positions: they are the *right* bargaining positions for individuals to have, not just those which individuals *happen* to have. The expanded moral backdrop sets a moral baseline. The no-agreement point therefore represents not the impossibility of justice given the circumstances, as it would in an unexpanded backdrop, but instead acts as justice's default position. In some cases this fact might affect what a person would consent to. Suppose that the moral backdrop includes libertarian self-ownership rights which may be replaced if their replacement passes the actual contractualist test. Suppose further that I am very much opposed to these rights because I see my role and that of others in my society as tools in the service of the glory of God. And suppose, finally, that everyone else in my society is also very religious, but

nevertheless affirms greater individual liberty than I do (if not as much as the libertarian moral backdrop). Now, against a minimal moral backdrop, I might have contemplated refusing to accept anything but my glory-of-God moral conception. It is highly unlikely that this moral conception could have secured unanimous acceptance, but in that case, if I had also refused to accept anything else, there would have been no moral conception that it would be just to impose on us all. Justice simply wouldn't have applied in our society, since the fallback position—the minimal moral backdrop—wouldn't have involved the imposition of any moral conception itself. I might have thought this preferable to giving what I see as a watered-down religious moral conception justice's seal of approval, and so have decided to hold out for my favoured moral conception or nothing at all. By contrast, against the libertarian moral backdrop, I might settle for the watered-down moral conception as morally closer to my favoured moral conception than the backdrop itself—better that justice involves some of my religious aims than that it involves none, as it will if there is no unanimity—and therefore accept it. So the fact that the expanded moral backdrop sets a moral baseline with justice's seal of approval may, quite apart from the fact that it sets (say) levels of material well-being and non-interference from others, inform the content of any moral conception that passes the actual contractualist test.

The moral backdrop, then, is a fundamental part of any actual contractualist theory. At minimum it includes the right not to be coerced to conform to some institutions and principles of justice that one does not actually accept, but it may be expanded to include any number of further rights, principles, and institutions. Such expansions are not inconsistent with the moral right which is the foundation of the theory's actual contractualism. Against the moral backdrop, entailed by it as necessary for the permissibility of any deviations from it, actual contractualists

specify a procedure of actual acceptance by which specified relevant individuals can come to be subject to a moral conception involving such deviations. As we have seen, this much is common to all actual contractualist theories.

Dual and mixed condition actual contractualist theories add a third component: an account of the content of the to-be-accepted parts of the advocated moral conception that is necessary, in addition to their actual acceptance-derived legitimacy, for them to be just. The idea here, as we saw in section 2 above, is that we have two dimensions of assessment of the relevant parts (which may be *all* the parts, as in dual condition actual contractualist theories) of moral conceptions—their legitimacy, understood in actual contractualist terms, and the independent soundness of their principles—in both of which the relevant parts must measure up in order to be counted, on the theory, just.

Now, this further regulative conception component in a dual or mixed condition actual contractualist theory of justice need not be connected in any way to the legitimacy component. One could consistently endorse brutal dictatorship, say, as necessary for justice even as one endorsed as also necessary a moral background and a legitimising process of actual acceptance which together made it incredibly unlikely by one's own admission that justice should ever be achieved (because no one would, against that background, actually accept the brutal dictatorship). But even setting aside the content of the regulative conception component (i.e. brutal dictatorship), this would be a rather unattractive view. What could motivate us to endorse both aspects of it at once? It is much more likely that a dual or mixed condition theorist would offer a regulative conception component which seemed likely at least in its normative implications (as opposed to the full argument for it) to gain the actual acceptance of the relevant individuals against the moral background

in question.⁹ The resultant view might still be curiously divided in the sense that there is no reason to suppose that it would come about *because* of the soundness of its principles (for all we have said the reasons that individuals would take themselves to have to actually accept it could have nothing to do with that), but it would not verge on the incoherent in the way that the brutal-dictatorship actual contractualism does.

This suggests the possibility of a theory whose favoured moral conception would be actually accepted by each relevant individual on precisely those grounds which are taken to justify it independently of its legitimacy according to its regulative conception component. (This, surely, is the political philosopher's ideal.)¹⁰ An actual contractualist theory like this would have a unity that is missing from those like the two I distinguished in the preceding paragraph. More significantly, it would include, as part of its account of why anyone should suppose that the relevant individuals would actually accept the moral conception, an appeal to *moral reasons* (those drawn upon in the regulative conception component) that those individuals are taken to have *and the fact that they take themselves to have them*. This would be, I think, of some interest in an actual contractualist theory. It is evident that nothing inherent in the actual contractualist interpretation of

⁹ This is how I read the relationship between Parts I and II of *Anarchy, State, and Utopia* (hereafter ASU) and Part III of the same book. In Part III Nozick offers arguments for a 'framework for utopia'. This framework turns out to be coextensive with the minimal state which he argues in Parts I and II is legitimate because of its compatibility with rights-respecting (largely) invisible hand processes which include actual acceptance. He openly disavows any connection between the legitimacy argument and the arguments for the framework at p. 333. See section 5 below.

¹⁰ See Michael Walzer's discussion of the 'heroic' political philosopher in "Philosophy and Democracy".

individualistic justification excludes appeal to the fact that individuals take themselves to have such reasons to accept (part of) a given moral conception in order to show why that conception would in fact actually be accepted. One tends to expect explanations only from self-interest or prudence of the actual acceptance of a given moral conception. Unified dual and mixed condition theories do not conform to this expectation. As I shall suggest in the next section, they are not as rare as one might suppose either: two of the most important contractualists in the history of philosophy set us the challenge of meeting at once both regulative conception and actual contractualist conditions of justice. But this aspect of their views tends to be underemphasised.

The introduction of moral reasons to explain individuals' acceptance of some moral conception should not be supposed to be especially objectionable. The key point is that (in addition to their role in the regulative conception component) they are introduced not as direct explanations of the legitimacy of the moral conception in question, but as explanations of individuals' actual acceptance of that conception which *itself* explains its legitimacy. The plausibility of an actual contractualist theory which unites its two components in this way turns, then, partly on the plausibility of attributing to the individuals whose actual acceptance of a proposed moral conception is in question acceptance of the arguments that constitute the regulative conception component. But this isn't especially objectionable.

This idea reaches its apotheosis in dual condition actual contractualism. On this view, every aspect of the advocated moral conception is subject to both a pure regulative conception test and a pure actual contractualist legitimacy test and must pass both in order for a society implementing it to be considered just. The explanation of its passing the latter test is explained in terms of each individual's acceptance of the moral conception on the basis of the reasons advanced in favour

of the former test. To the extent we find the ideal of individualistic justification appealing on its own and appropriately interpreted in the actual contractualist test, an actual contractualist legitimacy test which applies to the entire content of the favoured moral conception is to be preferred over one that doesn't. To the extent that it seems to us that good reasons for moral conceptions independent of anyone's acceptance of them suggests that there is something good about implementing them, a regulative conception test which applies to the entire content of the favoured moral conception is also to be preferred over one that doesn't. It may seem that these two justificatory ideas pull in different directions. But they can be reconciled in dual condition actual contractualism.

So far I have been outlining actual contractualist frameworks without relating them to familiar contractualist accounts. In the next section, then, I explain how Hobbes and Locke can be read as fitting into these frameworks. This will help to illustrate the foregoing discussion. In addition it will give us some reference points for a discussion of the appeal of actual contractualism, which I consider in the section after that.

5. Hobbes, Locke, Nozick, and actual contractualism

Both Hobbes and Locke can be read as actual contractualists in my sense. Their theories exhibit many of the features of the various types of actual contractualism that I have been describing. Each of them argues that through individuals' actual acceptance a moral conception may be instituted that would have no legitimacy otherwise, and each of them describes a moral backdrop against which this moral conception is established.

Hobbes's moral backdrop involves much less than Locke's, and in particular it involves no elements of a moral conception itself. Each person in the state of nature has what Hobbes calls a 'right of nature' which he defines as:

the liberty...to use his own power, as he will himself, for the preservation of his own nature...and consequently, of doing anything, which in his own judgment, and reason, he shall conceive to be the aptest means thereunto.¹¹

But Hobbes doesn't think that this right correlates with any duties on the part of others. It is a Hohfeldian liberty right. So in effect there are no norms of justice (in my sense) in the state of nature: nobody is under a coercively enforceable obligation to do anything. Although there are such things as the 'laws of nature' in Hobbes, most of these create no duties except when one is assured "that all shall observe the same laws towards him".¹² One can only be assured of this when there is "a common power, to keep [men] in awe"¹³—the sovereign. So the laws of nature create duties only after the social contract by which Hobbes supposes that a sovereign is created has been made.¹⁴

¹¹ Hobbes, *Leviathan*, I.14. He goes on to say that therefore "every man has a right to every thing: even to one another's body."

¹² *Leviathan*, I.15.

¹³ *Leviathan*, II.17.

¹⁴ The one exception might be Hobbes's fundamental law of nature: to seek peace (see *Leviathan*, I.14). But we can discount this, for (among other reasons) it is not a principle of justice in my sense, since it is not to be coercively enforced and it is not part of a moral backdrop which includes principles that are to be coercively enforced.

The Hobbesian contract itself consists in each individual's transferring his unlimited liberty right to the 'mortal god' which is the sovereign. This is normally effected through "the express words, *I authorise all his actions*".¹⁵ It would be consistent with Hobbes's framework for individuals to transfer only some portion of their unlimited rights to the sovereign.¹⁶ But Hobbes claims that "[t]he only way to erect...a common power" of the sort necessary to keep peace is "to confer *all* their power and strength upon one man, or upon one assembly of men, that may reduce all their wills...unto one will".¹⁷

This brings us to individuals' reasons for making the contract by which civil society is established. Famously, the state of nature is, in Hobbes's view, a "condition of war of every one against every one". Life in it is "solitary, poor, nasty, brutish, and short".¹⁸ Escape from this "miserable condition" into a state in which the awe of a common power "tie[s] them by fear of punishment to the performance of their covenants, and observation of [the] laws of nature" is the point of the contract.¹⁹ So is Hobbes here appealing to individuals' acceptance of moral reasons for the moral conception that he advocates, or to mere self-interest? Both types of explanation are present in *Leviathan* and, correspondingly, Hobbes can be read both as advocating a dual condition and as advocating (almost) a pure legitimacy theory.²⁰ The former insofar as the laws of nature and other features of the Hobbesian moral conception represent necessary conditions of the justice of that moral conception,

¹⁵ *Leviathan*, II.21.

¹⁶ And indeed the right to defend one's own body cannot be transferred, as I noted above.

¹⁷ *Leviathan*, II.17, my italics.

¹⁸ *Leviathan*, I.13.

¹⁹ *Leviathan*, II.17.

²⁰ The qualification 'almost' has to do with the inalienability of the right to defend oneself.

and the latter insofar as it is advocated solely on the grounds that it is what individuals do accept. Insofar as Hobbes's is a dual condition theory, it displays the interesting unity that I described in the preceding section: the reasons that Hobbes takes there to be for making the moral conception he advocates independently necessary conditions on the justice of society are also those that he takes individuals to affirm and consequently upon which to base their acceptance of the moral conception that the contract licenses.

Hobbes's would be a *dual* condition rather than a *mixed* condition actual contractualism—or, alternatively, it would be a *pure* legitimacy theory rather than a combined theory of any sort—because of the normative emptiness of the moral backdrop that I described above. Setting aside the inalienable right to self-defence (which is nevertheless consistent with the sovereign's punishment of one), no aspect of Hobbes's favoured moral conception is just unless it has been accepted by each individual subjected to it. (An individual is not subject to it who withholds his acceptance.)²¹ In a mixed as opposed to a dual condition theory, recall, some parts of the advocated moral conception are either not subject to the actual contractualist test or else not subject to the regulative conception test. Insofar as Hobbes can be read as offering both tests as conditions of the justice of a moral conception, it is plausible to suppose that they both take in the full extent of that conception. In any case it is plausible to read him as subjecting all parts of his favoured moral conception to at least the actual contractualist test.

This is a respect in which Hobbes contrasts with Locke. On Locke's account in the *Second Treatise of Government*, the legitimacy of the state's authority depends upon

²¹ See *Leviathan*, II.18.

the actual consent of those subject to it.²² But Locke's moral backdrop includes what is, in my terms, a moral conception. It includes principles permitting the coercive enforcement of a range of principles pertaining to the distribution of liberties and goods. These principles would remain in force if no one ever exited the state of nature via the social contract.

The institutions and principles of Locke's moral backdrop are not all absolute; a moral conception may involve institutions or principles which contradict some of them and still be just if it has been actually accepted by each person subject to it. (No one, however, may "willfully" give up his own life to an arbitrary power.)²³ As we saw above, even non-absolute moral conceptions in the moral backdrop may inform the content of the moral conception ultimately adopted; and so it is with Locke. Hobbes's moral conception includes a certain equality of treatment of each member of society (see e.g. *Leviathan*, I.15) which, if he is interpreted as a pure legitimacy theorist, is best explained by the effective equality of bargaining positions in his miserable state of nature. The life of even the strongest remains solitary, poor, nasty, brutish, and short.²⁴ And to the extent that Locke's favoured moral conception subjects the equal rights and liberties which it grants individuals to the actual contractualist test, these too are to be explained by the equality of individuals' bargaining positions in the state of nature. But a better explanation of Locke's endorsement of this equality is his *moral* backdrop, the equality of (libertarian) rights that each person has in the state of nature.

²² Locke, *Two Treatises*, II, §95.

²³ *Two Treatises*, II, §6 and §23.

²⁴ See Allan Gibbard, "Constructing Justice", p. 270.

To explain. Locke, like Hobbes, does mention the insecurity of life in the state of nature. But he places no weight on the idea that this arises through disobedience to the law of nature. Instead he emphasises the individual insecurity inherent in a situation in which each has the right to judge and enforce the law of nature in his own case, as he supposes we do in the state of nature.²⁵ People are, as they are in Hobbes's state of nature, "constantly exposed to the invasion of others".²⁶ But this is because of partiality in the administering of natural law rather than because of Hobbesian licence.²⁷ This being so, it is plausible to suppose that for Locke, the contract is to be made under the conditions of individual freedom created by widespread compliance with the natural law in the first place, rather than under conditions in which retributive violence has reached such a level that the situation is indistinguishable from a Hobbesian state of war. (Indeed, he emphasises that the state-founding consent must be free, which implies that at the time of each individual's consent, at any rate, his natural liberties were being respected; and he denies at one point that "men are bound by promises, which unlawful force extorts from them.")²⁸ Arguably, were it not made under these conditions, it would not legitimise the government it creates.

The Lockean law of nature, unlike the law of the state established through the contract, is not subject to the actual contractualist test. Since it informs the content

²⁵ See especially *Two Treatises*, II, §§123-7. Also §90: "for the end of civil society being to avoid and remedy those inconveniences of the state of nature, which necessarily follow from every man's being judge in his own case..."

²⁶ *Two Treatises*, II, §123.

²⁷ Cf. *Two Treatises*, II, §6: "though this be a state of liberty, yet it is not a state of licence."

²⁸ See *Two Treatises*, II, §§192, 176. Cf. Rawls, *Theory*, pp. 112-3/96-7 (original version / revised version).

of the moral conception ultimately advocated by Locke—because acceptance must be made under conditions of widespread compliance with it—there is a sense in which Locke's actual contractualism is less thoroughgoing than Hobbes's.

Although in theory—to the extent that Locke is a pure legitimacy theorist—any moral conception might be accepted by individuals in Locke's state of nature and consequently justified, as a matter of fact it seems highly unlikely that at least a large number of individuals will accept anything which restricts their liberties further than the natural law does, as indeed Locke points out at *Two Treatises*, II, §131.

It is possible to read Locke as a mixed condition actual contractualist. (Not a dual condition actual contractualist since it is possible on Locke's view to have a justified moral conception—and thus, in my terms, a just society—which is not accepted by any individual. That moral conception is of course that of the law of nature.) He makes repeated reference to the features and aims of a just state (chiefly the protection and preservation of individuals' property, understood to include "lives, liberties, and estates").²⁹ These features, though often invoked as reasons for individuals to accept such a state, sometimes also seem to be viewed by Locke as necessary conditions themselves.³⁰ So it may seem that, for Locke, a government which does not act in accordance with that primary aim, or which violates the constraints in question, cannot have the authority it claims—even if the members of society have consented to this authority. To the extent that Locke really is committed to this view, he is in my terms a mixed condition actual contractualist:

²⁹ *Two Treatises*, II, §123.

³⁰ See for example *Two Treatises*, II, §§131, 135, 138, 142.

the regulative conception test is applied to all aspects of just moral conceptions, the actual contractualist test to some.³¹

What of the process of actual acceptance itself on a Lockean picture? Like Hobbes, Locke founds the state on express consent. Tacit consent famously plays a role, but it is secondary. For although Locke does discuss tacit consent, he does so in the context of the problem of “how far anyone shall be looked on to have consented...where he has made no expressions of it at all”.³² In reply, he says that when someone “by inheritance, purchase, permission, or otherways, enjoys any part of the land” governed by the state, he submits to its authority—but only for *as long as* he enjoys any part of the land.³³ Since, presumably, there is at the time of the original contract no state and so no possibility of enjoyment of land under the jurisdiction of the state to count as tacit consent to the state’s authority, and since Locke describes tacit consent in no other context, it seems to me that the state can come to govern that land in the first place, on Locke’s view, only through express consent. Thus express consent is the primary foundation of the state’s legitimate authority over individuals.

Free individuals could in principle consent to any moral conception apart from one which involves their willingly giving up their lives to an arbitrary power, but

³¹ Cf. Simmons, “Justification and Legitimacy”, pp. 745-6. Simmons thinks that as well as (in my terms) an actual contractualist account of individual states’ legitimacy, Locke offers a “general argument for the moral and prudential preferability of states ruled by limited governments to life in the state of nature”. An argument for the *moral* preferability of such states certainly amounts to what I am calling a regulative conception test (albeit a somewhat limited one, since it determines only that one type of arrangement is preferable to one other) in Locke’s view.

³² *Two Treatises*, II, §119.

³³ See *Two Treatises*, II, §120-21.

Locke emphasises the rationality of their consenting only to a moral conception which safeguards their property.³⁴ As with Hobbes, this can be read in two ways. To the extent that Locke is a mixed condition theorist, we can understand his theory as having some of the interesting unity whereby the reasons given in the regulative conception conditions on a just society are also those for which it is supposed that individuals will actually accept one's favoured moral conception and so legitimise it according to the pure legitimacy component. This helps, in addition, to make Locke's view more plausible in the following sense. As we saw a moment ago, Locke's theory is plausibly read as starting from a state of nature in which there is widespread compliance with the law of nature. If, however, individuals in that state of nature are understood to be entirely self-interested to the point of disregard for the law of nature, then Locke's theory may be correct but will appear far more utopian than he seems to have taken it to be. For why should even a committed Lockean suppose that we will ever have the conditions under which actual acceptance of Locke's favoured moral conception will legitimise it? But if individuals in the state of nature are instead to be understood as broadly law of nature-abiding, then this problem (*qua* internal difficulty in Locke) dissolves.³⁵

To the extent that Locke's is not a mixed condition theory but a pure legitimacy theory, the passages emphasising the rationality of consent to a moral conception which safeguards individual property must be read as full explanations of why *this* (our, Locke's) moral conception ought to have the aims and features in question, *viz.*, because those aims and features are what was as a matter of fact consented to.

³⁴ See for example *Two Treatises*, II, §§21, 120, 123-6, 131, 136, 149.

³⁵ Also pertinent in this regard is Locke's restriction of scope to those "capable of knowing the law". See *Two Treatises*, II, §§57-63.

If some other aims and features had been consented to, this state would have been unjust to have these aims and features instead of those. The rationality of consent to this moral conception can be seen as a guide to unverifiable historical fact.

Whichever interpretation we adopt of either Locke's or Hobbes's view, we should be careful not to lose sight of the fact that historical fact, and not any putative guide to it such as rationality, has precedence. If we were to discover that individuals, even irrationally, consented in fact to some moral conception other than that which Hobbes or Locke favours, this would make the imposition of the latter unjust. On pure legitimacy readings it would also be sufficient to make the imposition of the former *just*. The fact that it would not have been rational to choose this way cannot, on an actual contractualist theory, be sufficient to render such a choice irrelevant to the justice of a moral conception. On an actual contractualist theory, actual acceptance is a necessary condition of justice: that is the point.

But as I argued in Section 3 above, there is some room for manoeuvre with respect to whose actual acceptance counts and what counts as their actual acceptance. If one restricts scope to those who are rational, say, one may expect that the appeal to rationality as a guide to historical fact will be more accurate. Still, although this rules out the irrational choice of an irrational individual as pertinent to the justice of a moral conception, it does not rule out the irrational choice of a rational individual. If no gap is left between what someone must (rationally) accept and what she does (actually) accept, we are no longer actual contractualists, since we are no longer appealing to any real exercise of individual agency.

Hobbes appears not to restrict scope even to those who are rational. His arguments for the moral conception he advocates turn ultimately on the need for or law of self-preservation, which is plausibly enough attributed to every individual.

Since what counts for Hobbes is the will of these contracting parties, and as he makes clear at *Leviathan*, I.6, voluntariness is independent of “rational appetite”, there appear to be no further restrictions. Locke is more problematic. He does restrict scope, excluding those not capable of “such a degree of reason, wherein [they] might be supposed capable of knowing the law”.³⁶ And he takes having this degree of reason to amount to being someone who chooses in her “proper interest”.³⁷ So to preserve Locke as any kind of actual contractualist at all we had better place weight on the only individuals he mentions as excluded by his definition, namely children and “*lunatics* and *idiots*” rather than anyone who might choose other than in her self-interest.³⁸ Otherwise the free consent of contracting parties cannot be other than as Locke specifies; and failure to consent, too, would automatically rule one out of the group of those whose consent is necessary to legitimise the state. I shall therefore assume that it is consistent with Locke’s specification of scope that the moral conception that meets the test of actual acceptance might at least in principle be other than that which Locke describes as proper to the state. (This is not an unorthodox reading.)

One other actual contractualist whose views it is worth describing by way of illustration of actual contractualism is Locke’s contemporary descendant Robert Nozick. Nozick is interesting from the point of view of my current discussion because he clearly delineates the various features of actual contractualism (as I term it) in his argument for his favoured moral conception, the ‘minimal state’. Like Locke, Nozick envisages a state of nature regulated by a deontological moral

³⁶ *Two Treatises*, II, §60. By ‘the law’ Locke means the law of the state of nature. People not capable of such a degree of reason are not free in an important sense. See §§56-63 and below.

³⁷ See especially *Two Treatises*, II, §§57-9.

³⁸ *Two Treatises*, II, §60.

conception of ‘rights as side constraints’, which he explicitly affirms as the moral backdrop for the actual contractualist legitimising transition to the minimal state. (Though unlike Locke he sees all these rights as alienable.)³⁹ Like Locke, Nozick restricts scope fairly drastically, but he presents an argument for doing so, claiming that this moral conception is justified as the moral backdrop because of an inviolability that humans ought to have which is founded on their rationality, among other things.⁴⁰ (Nozick is less insistent than Locke, however, that this should inevitably lead to the acceptance of the moral conception that he favours: a peculiarity of his account is that at times he suggests that it merely needs to be plausible that this conception *could have* met the actual contractualist test for it to be just.)⁴¹ Like Locke, Nozick goes on to argue that against this moral backdrop his favoured moral conception would arise through voluntary transactions between individuals and, later, between individuals and ‘protective agencies’;⁴² but unlike Locke, he is explicit about his conception of voluntariness.⁴³

³⁹ See e.g. *ASU*, p. 331. On occasion Nozick retreats from his commitment to the alienability of our rights. See e.g. pp. 292-3, where he says that *only* an ‘analogue’ of a more-than-minimal state under which each person retained the option to “opt out” of its arrangements could be legitimate.

⁴⁰ I say more about this in Chapter 6.

⁴¹ See for example *ASU*, p. 276; cf. Nozick’s discussion of ‘explanatory political theory’ at pp. 6-9. Simmons reads Nozick as simply having two aims in mind: first, to show that a minimal state could arise and would be better than anarchy; and second, to show under what conditions any particular state is legitimate. See Simmons, “Justification and Legitimacy”, p. 744. Cf. Rawls, *Lectures on the History of Political Philosophy*, pp. 30-2.

⁴² Voluntary transactions involving *some* members of society under the ultimately accepted moral conception are necessary, but not involving *all* of them: for the minimal state rules legitimately even vis-à-vis some within its territory who have not participated in such transactions. This is because of the prior right of a would-be state to prohibit non-participants from enforcing their rights themselves

Nozick's view also illustrates very clearly the distinction between the actual contractualist legitimacy condition and the regulative conception condition. Insofar as a regulative conception argument informs the moral backdrop of an actual contractualist theory, its relation to the legitimacy condition is somewhat complex: as we have seen, it determines the content of one possible just moral conception (that of the state of nature) and at the same time may inform but does not necessarily determine the content of another just moral conception, the one favoured by the theorist. Things are simpler when the regulative conception argument operates entirely independently as it does (to the extent that there he offers such an argument) with Hobbes.

Nozick offers both the backdrop-determining version and the independent version. We find the latter in Part III of *Anarchy, State, and Utopia* (hereafter *ASU*). There Nozick offers arguments for a 'framework for utopia' which turns out to be coextensive with the minimal state, and which in Parts I and II he has argued is justified because of its resulting from actual contractualist acceptance against the

so long as it compensates them by enforcing them on their behalf (see *ASU*, chapter 5). So Nozick's is a split condition theory to that extent: some aspects of the moral conception he endorses are just although they do not meet the actual contractualist test; others are just only because they meet that test. In fact the overall view is a mixed theory: the moral backdrop includes coercively enforceable principles which explicitly inform—Nozick makes a virtue of this—the moral conception that would be accepted, as with Locke; and as I go on to point out in the main text, Nozick also offers an entirely independent regulative conception argument for the minimal state.

⁴³ Nozick goes beyond Locke in claiming that only by violating rights could a state more extensive than his favoured minimal state arise (e.g. at *ASU*, p. 333). But this doesn't seem consistent with his belief in the alienability of our rights, or with the "eldritch tale" which he uses to explain how "what is recognizable as a modern state...might arise from a minimal state without any blatant violation of anyone's rights through a series of individual steps each arguably unobjectionable" (p. 290).

moral backdrop of rights as side constraints. Nozick openly disavows any connection between this latter justification and Part III's arguments for the framework for utopia.⁴⁴ If Nozick supposes that the argument of Part III justifies the minimal state and shows every other moral conception to be unjustified, then *ASU* as a whole offers an interesting mixed condition actual contractualism. Actual acceptance is, as we see from Parts I and II, necessary for the justice of the minimal state. It is informed by, and to some extent determined by the (also necessary) moral backdrop which Nozick supports with a regulative conception argument. But there is also the independent regulative conception argument of Part III which imposes a further necessary condition on the justice of imposing a moral conception. It must be a framework for utopia. Individuals' acceptance against the appropriate moral backdrop of some other moral conception would be insufficient for its justice.

(This mixed condition theory would not have the interesting unity that I discussed earlier. Nowhere does Nozick suggest that individuals' legitimising acceptance of the minimal state would be based on reasons furnished by the arguments of Part III.)

This brings to a close my exposition of the structure of actual contractualism. For the remainder of this chapter, I focus on the underlying motivation for actual contractualism.

⁴⁴ At *ASU*, p. 333.

6. *Autonomy and the appeal of actual contractualism*

As we have seen, it is a fundamental presupposition of actual contractualism that individuals have a prior moral right against (at least) coercion to behave in accordance with (at least some) institutions and principles insofar as they do not accept those principles. I have already said that the actual contractualist interpretation of the contractualist ideal of individualistic justification is attractive. But why is it so attractive? What is it about being able to say '*you* accepted these terms' to each person that we care about? An obvious and standard answer is that it is a concern for the value of individual autonomy. In this section I want to argue that at least in the form that might be thought to underlie the actual contractualism of such theorists as Locke or Nozick, this value can't be the whole story of the attraction of actual contractualism, since it fails to explain why Locke's is not *clearly* a superior theory compared with Hobbes's. Instead I'll suggest a more abstract conception of autonomy as the source of actual contractualism's basic appeal. More will need to be said, however, to explain how what I called 'thoroughgoing' actual contractualism taps this source. Further on, in Chapter 3, I argue that Rawls offers us a plausible explanation of the right sort.

In one familiar sense, individual autonomy seems fairly clearly to explain at least part of the attraction of some actual contractualist views. This is the sense that we tend to have in mind when we talk of living an autonomous life, of bringing a person up to be autonomous. (I'll use 'autonomous living' to mean this sense of individual autonomy.) Actual contractualists, as we have seen, arm each person with the power *actually* to veto putative conceptions of justice. A principle of justice does not meet the contractualist ideal of individualistic justification, and so a society's moral conception is not just, if any one person in that society does not

actually accept it.⁴⁵ This veto is granted to each person in the form of the moral right that I introduced earlier against, minimally, coercion to behave in accordance with some institutions and principles of justice which she does not actually accept. In some cases, this right can be seen as one element in the service of the ideal of autonomous living. The basic thought is as follows. Autonomous living requires that the way in which one's life goes, the things one does and thinks, and the obligations that one takes on, are controlled to at least a fairly large extent by oneself. You do not achieve the aim of autonomous living if others or circumstances alone determine what you do and think and the obligations that you take on. So a theory which aims at the ideal of autonomous living for individuals must make it the case that the things a person does and thinks are open only or ultimately to her control, or else open to others' control only because she has accepted, in some sense, that they should have it. Hence a range of rights against aggression and force and the like. Similarly, we might suppose, her obligations should be taken on by her rather than imposed on her from without. Hence the actual contractualist right. Note that the rights against aggression etc., however, could be imposed without anyone's actual acceptance, since they exist to serve the same ideal that the actual contractualist right serves. These would exist alongside that right even prior to anyone's actual acceptance of anything.

This description fits those actual contractualist theories whose moral backdrop—against which the transition to the legitimate imposition of a moral conception takes place via actual acceptance—includes more than the bare actual contractualist right, including those of Locke and Nozick. The moral backdrop for

⁴⁵ Although a conception of justice may meet the individualistic ideal with respect to everyone who does accept it, it fails to do so with respect to a society as a whole in which not everyone accepts it.

their actual contractualism includes equal, robust rights of self-ownership for all, granting each person the coercively enforceable right, in Locke's words, "to dispose, and order as he lists, his person, actions, possessions, and his whole property...and therein not to be subject to the arbitrary will of another, but freely follow his own."⁴⁶ Although one might argue that this conception of autonomous living is too minimal, so that it fails genuinely to guarantee each person the possibility of an autonomous life, it seems reasonable nevertheless to claim that it *is* a conception of autonomous living rather than anything else that drives this specification of the moral backdrop and the actual contractualist condition on changes to it. The Lockean ideal of autonomous living may be more akin to Berlin's negative liberty than to a substantial Razzian ideal, but it nevertheless makes sense to suppose that a fundamental Lockean commitment is that individuals live autonomously in the state of nature.

But neither Locke nor Nozick is what I called a 'thoroughgoing' actual contractualist. The moral backdrops that they specify are influential even if not unalterable (that depends upon whether individuals may through acceptance legitimise a less libertarian moral conception),⁴⁷ and they are not subject to the actual contractualist test. So to an extent the moral conceptions ultimately endorsed by Locke and Nozick reflect a set of norms which do not satisfy the ideal of individualistic justification as it is interpreted by actual contractualists. Perhaps this might be taken merely to reflect the implausibility of taking that ideal to require anything more were it not for the fact that we have, in Hobbes, a thoroughgoing

⁴⁶ *Two Treatises*, II, §57. Nozick approvingly quotes similar passages from Locke. Clearly Locke does not mean to include the freedom to dispose of these things in ways that violate the duties that others' having the same freedom places on each person.

⁴⁷ See section 4 above.

actual contractualist whose theory is not *ipso facto* implausible. Hobbes, as we saw, specifies (at least on one reading) a moral backdrop which is normatively empty apart from the requirement of individual acceptance to legitimise a moral conception.

Now I think that it is wrong to say that Hobbes's theory is driven (or that the attraction derives from), as Locke's and Nozick's may plausibly be thought to be, the ideal of autonomous living. I think this for the following reasons. On even a fairly minimal conception of what the conditions for an autonomous life involve, an autonomous life is surely one whose direction is not predominantly governed by the need to survive.⁴⁸ So we can say that individuals live autonomously only in conditions in which their lives are not predominantly governed by the need to survive. If an actual contractualist theory is based on the ideal of autonomous living, then, its moral backdrop should be such that individuals in it should not—normally, at any rate—lead lives which are predominantly governed by the need to survive.

But in Hobbes's state of nature, individuals' lives *are* normally predominantly governed by the need to survive. That they are 'nasty, brutish, and short' is due to the difficulty of meeting this need: "there is no place for industry; because the fruit thereof is uncertain...and which is worst of all, continual fear, and danger of violent death".⁴⁹ So either it should not be thought of as based on autonomous living (in even a minimal sense) or else it is, but fails to realise that value adequately. If the latter were the case, there would surely be very little to recommend a Hobbesian view—Locke's theory would be *clearly* an improvement. But I take it that Hobbes's

⁴⁸ Cf. Raz's 'hounded woman' example at *The Morality of Freedom*, p. 374.

⁴⁹ *Leviathan*, I.13.

theory is not plainly an inadequate or underdeveloped version of Locke's. And part of the explanation of its attraction is the thoroughgoing nature of its actual contractualism.

One objection to this analysis of Hobbes is that it casts doubt on the idea that even Locke's view could be based on autonomous living. For one might suppose that in the sense that they may be locked into a spiral of escalating retribution for rights violations, individuals in Locke's state of nature too lead lives predominantly governed by the need to survive. But it is plainly much more plausible to suppose that Locke's view is based on the value of autonomous living. So the argument against reading Hobbes as committed to the same value must be mistaken.

But this is to ignore the point I made in the preceding section that Locke should be taken to view actual acceptance as legitimising the accepted moral conception only against a background of widespread compliance with the laws of nature. It is this background which he should be understood to endorse as a state of autonomous living, even if it carries with it the *risk* of the sort of (Lockean, rather than Hobbesian) state of war that makes it rational for individuals to accept a moral conception which deprives them of individual rights of punishment. If Locke supposes that acceptance in a situation where retributive violence had reached a bloody peak could legitimise the accepted moral conception, it would be plausible to doubt that his view was aimed at autonomous living too. But he does not.

What, then, explains the attraction of Hobbes's thoroughgoing actual contractualism? One suggestion is that it is radically constructivist, an explanation of the foundation of morals: *all* morals are to be constructed through the legitimising process of actual acceptance against a backdrop of *no* morals. (This is

how David Gauthier's Hobbesian approach works.)⁵⁰ But this doesn't sit easily with the interpretation of Hobbes as a *dual condition* actual contractualist. Hobbes, I have claimed, supposes that his favoured moral conception can be "found out by reason" independently of anyone's actual acceptance, and thus can be read as including a regulative conception component in his view. As I read his view, it's therefore not the case that just any moral conception can be legitimised by actual acceptance, as it would be on a radically constructivist view which privileged actual acceptance as Hobbes does. Nevertheless, actual acceptance too is a necessary condition of the existence of the obligations of the favoured moral conception.

So the ideal of autonomous living is not, I conclude, at the heart of actual contractualism's appeal—not in all its forms, at any rate. Could there be some other explanation which would account for the appeal of more thoroughgoing actual contractualism? In Chapter 3, drawing on Rawls, I shall suggest that a conception of the individuals as epistemic agents capable of reaching different but equally reasonable conclusions about the moral world makes sense of the appeal of a more thoroughgoing actual contractualism. First, though, I turn to what I call 'modal contractualism', the rival to actual contractualism as an interpretation of the individualistic ideal.

⁵⁰ See Gauthier, *Morals by Agreement*.

Chapter 2: Modal Contractualism and Hybrid Contractualism

1. Modal contractualism	68
2. What's distinctive about modal contractualism? – the question	71
3. What's distinctive about modal contractualism? – the answers	75
4. First-order normative modal contractualism	89
5. Modal contractualism and actual contractualism	94
6. Two puzzles for hybrid theorists	99
7. Conclusion	108

1. *Modal contractualism*

The ideal of individualistic justification that underlies contractualism is subscribed to not only by actual contractualists but also by those whom I shall call ‘modal contractualists’. But modal contractualists interpret the ideal in way that is very different from the actual contractualist interpretations. According to modal contractualists, a moral conception might satisfy the ideal even if no one in the relevant society actually accepts it. For modal contractualists think that what matters is whether or not anyone *could reasonably* reject it. What individuals actually accept and reject is irrelevant: they might unreasonably actually reject a moral conception, or they might actually accept one even though they could reasonably reject it. In this section I describe and analyse modal contractualism.

I focus on the views of T.M. Scanlon and Thomas Nagel, who are the pre-eminent contemporary exponents of modal contractualism. The term is mine, however, so before I go on I should say something to justify the introduction of a

new term to describe their views. I note first that Scanlon himself describes his view simply as contractualist. But plainly his view is not what I have been calling *actual* contractualist. Nagel describes Scanlon's view, and by implication his own similar view, as 'hypothetical contractualism'.⁵¹ But that will not do for our purposes here for the following reasons. First, Scanlon straightforwardly denies that his view is hypothetical contractualist, and I think that he is right to do so.⁵² On a natural understanding of hypothetical contractualism, he argues, claims about moral wrongness are identified with non-normative claims about what people *would* or *would not* do under certain conditions. This is how Rawls is most straightforwardly read, and his theory has been viewed as the paradigm of a hypothetical contractualist theory. Scanlon does not identify claims about moral wrongness with non-normative claims in this way; on the contrary, he says, he identifies them with claims about what individuals could reasonably reject which are unambiguously normative themselves. So even if it is true to say that Scanlon's theory deals with hypothetical questions, to use the same term for both Scanlonian and Rawlsian contractualist views associates them by definition when in fact the claim that one should be assimilated to the other has been regarded as a philosophical novelty, not a trivial truth.⁵³ So it makes sense to use different terms for their different views.

A further reason pertains specifically to this thesis. I want to argue that Rawls should not be read, as he often has been since Scanlon published "Contractualism and Utilitarianism", as advancing a view which straightforwardly identifies claims

⁵¹ See Nagel, *Equality and Partiality*, p. 37.

⁵² See Scanlon, "Replies", pp. 434-5.

⁵³ See Scanlon, "Contractualism and Utilitarianism".

about what is wrong (or unjust) with claims about what no one could reasonably reject. Instead, his hypothetical contract theory should be seen as combining both the Scanlonian approach and the actual contractualist framework that I described in Chapter 1. To describe both Rawlsian and Scanlonian forms of contractualism with the same term would tend to obscure the structure of my argument. So, again, it makes sense to describe them using different terms. Given the established association of Rawls with hypothetical contractualism and Scanlon's renunciation of this label to describe his own view, it seems to me permissible to introduce a new term for the latter. 'Modal contractualist' seems no worse than any other, so that is what I shall use.

The essence of modal contractualism is articulated by Scanlon's contractualist formula:

An act is wrong if and only if any principle that permitted it could reasonably be rejected by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject.⁵⁴

This formula covers the full range of 'what we owe to each other', but as Nagel notes it is natural also to consider it only in application to the special case of justice.⁵⁵ Thus we can say that the following articulates the essence of modal contractualism as it applies to the case we are interested in:

⁵⁴ This formulation paraphrases what Scanlon says at p. 4 of *What We Owe to Each Other* (hereafter *WWO*).

⁵⁵ See Nagel, *Equality and Partiality* (hereafter *EP*), pp. 36-7. Scanlon seems to agree that political philosophy might put his contractualism to work in the sphere of justice specifically. See *WWO*, p. 245.

A moral conception is just (or at least not unjust) if no one could reasonably reject it as a moral conception for the regulation of their single society given the aim of finding a moral conception which no one could reasonably reject who was similarly motivated.

I shall move between these two formulations as seems appropriate in the context. But it makes no difference to my argument which of them we have in mind as we discuss modal contractualism and actual contractualism.

2. *What's distinctive about modal contractualism? – the question*

What does it mean to say that no one could reasonably reject a principle or moral conception? It means, to begin with, that the strongest applicable reasons against that principle or moral conception are weaker than the strongest applicable reasons against any competing principle or moral conception. It does not mean that there are *no* reasons against it; but that despite any such reasons that there may be, it would not be reasonable for anyone to reject it given the stronger reasons that there would be for rejection of any principle or moral conception that could take its place.

It's plain even from this incomplete description that modal contractualism exhibits the following feature which gives it an attraction to rival those of actual contractualism: it is concerned with justifiability *simpliciter* rather than *apparent* justifiability. The actual contractualist interpretation of the individualistic ideal, by contrast, is concerned with apparent justifiability rather than justifiability *simpliciter* in the sense that what matters for actual contractualists is that each person *takes* a given moral conception to be justified, as is expressed by her actual acceptance. Of

course actual contractualists consider universally actually accepted moral conceptions to be justifiable too. But this flows from the fact that individuals take them to be so. Modal contractualists do not, at the fundamental level, concern themselves with apparent justifiability, the question whether anyone *takes* a given moral conception to be justified, at all.

They are, however, inspired by the idea of the individual veto which is realised by actual contractualists. As we have seen, actual contractualists require unanimity for legitimacy and hence justice. Modal contractualists also aim at unanimity;⁵⁶ but rather than the actual unanimity familiar from actual contractualist views, they aim at a kind of idealised unanimity, the unanimity of reasonable individuals.

Yet if the relevant unanimity of reasonable individuals on a moral conception is obtained if and only if no one could reasonably reject that moral conception, and this is explained in terms of its being supported by the strongest reasons, we might wonder whether modal contractualism is really *contractualist* at all. We might wonder, that is, whether it genuinely offers an interpretation of the ideal of individualistic justification. For how, we might ask, is the view that a moral conception is just only if no one could reasonably reject it who was motivated to find principles that no one could reasonably reject who was similarly motivated substantively different from a ‘straightforward view’ which “dispens[es] with the contractualist packaging”⁵⁷ and says simply that a moral conception is just only if it is best

⁵⁶ Nagel explicitly invokes unanimity as his guiding idea. See *EP*, pp. 33ff; also “Equality”, pp. 116-25. Scanlon tends to talk of ‘general agreement’ rather than unanimity (see e.g. *WWO*, p. 5), but he means by general agreement the agreement of *all* parties rather than the agreement of most. (Scanlon invokes Nagel’s discussion and endorsement of unanimity rather than majority as the guiding idea at p. 144, n. 17 of “Contractualism and Utilitarianism”.)

⁵⁷ Scanlon, “Replies”, p. 434.

supported by the applicable reasons that exist? I take it that there is nothing obviously individualistic about this latter view. It doesn't involve any 'one-by-one' justification (see the Introduction, above). But what does framing the issue in terms of idealised unanimity add?

The worry here can be brought out by considering arguments that people might make for and against utilitarianism. You might, employing the modal contractualist terminology, say: 'no one who was motivated to find principles which no one could reasonably reject who was similarly motivated could reject utilitarianism, because it maximises utility.' And I, employing the same terminology, might reply: 'on the contrary, utilitarianism could reasonably be rejected by those with the appropriate motivations, for it favours a situation in which many people are discomfited for ten minutes and one person is in extreme and enduring pain over a situation in which everyone is discomfited for an hour and no one is in extreme pain at all—so long as the many are sufficiently many that their extra fifty minutes of discomfort in the second situation add up to more disutility than that experienced by the person in extreme, enduring pain. And that is grounds for rejection because it violates Parfit's *Priority View*, according to which "benefiting people matters more the worse off the people are to whom these benefits would go."⁵⁸ The many would have to number more than utilitarianism supposes they would number before it would be reasonable to favour the first situation over the second.⁵⁹ But couching our arguments in

⁵⁸ Parfit, "Justifiability to Each Person", p. 382.

⁵⁹ I employ this rather complex reply rather than the more straightforward anti-aggregative claim that follows from Scanlon's 'Individualistic Restriction' because as I shall argue shortly, the Individualist Restriction and its anti-aggregative consequences are precisely what *do* distinguish the modal contractualist idealised unanimity view from the straightforward view that what's just is what's best supported by the reasons that there are.

modal contractualist terms in this way doesn't seem to add anything to the substance of our claims. The force of your defence and the force of my reply seem completely unchanged if you claim simply that there is good reason to maximise utility and so to adopt utilitarianism and I claim simply that the Priority View gives us good reason not to adopt utilitarianism.

Actual contractualists, then, by giving each person an actual veto over putative institutions and principles of justice, offer us a clear sense in which actual contractualism interprets the individualistic ideal. But it's not immediately obvious that we can say the same of modal contractualism. So is there really anything distinctively contractualist about modal contractualism? In the next section I consider some candidates that modal contractualists have suggested to distinguish modal contractualism from the straightforward view. Unless there are such distinguishing features, then modal contractualism isn't contractualist, since the straightforward view certainly isn't. First I consider, and reject, the idea that the modal contractualist account of motivation distinguishes it. Then I consider the idea that the identity of the group of people that we mean when we say that no one (none of them) could reasonably reject a just moral conception makes a difference. I reject this idea too. Finally, I consider the idea that modal contractualism's incorporation of the 'Individualist Restriction'—according to which a condition of modal contractualist reasonable rejectability is that the grounds for rejection are elaborated in terms of the implications only for single individuals—distinguishes it from the straightforward view. This, I conclude, is a viable candidate. I therefore take modal contractualism to be distinguished by its incorporation of the Individualist Restriction.

3. *What's distinctive about modal contractualism? – the answers*

Nagel offers the following elaboration of modal contractualism's distinctive interpretation of the contractualist ideal:

As I have said, the search for legitimacy is a search for unanimity... The unanimity in question is neither actual unanimity among persons with the motives they happen to have, nor the kind of ideal unanimity that simply follows from there being a single right answer which everyone ought to accept because it is independently right, but rather something in between: a unanimity which could be achieved among persons in many respects as they are, provided they were also reasonable and committed within reason to modifying their claims, requirements, and motives in a direction which makes a common framework of justification possible. [...]

If such a hypothetical unanimity were discoverable, it would explain the rightness of the answer rather than being explained by it. That is, it would not be possible to discover what everyone should agree to by following a single course of reasoning which everyone can follow to the right result. Rather, the right result would be discoverable only by finding that different persons, reasoning from their different perspectives, will converge on it.⁶⁰

It seems to me that we can discern in this passage two suggestions as to what is distinctive about modal contractualism. The first is that the reasonableness of those whose perspectives are being taken into account makes a difference to which principles could be reasonably rejected. What people 'provided they were also

⁶⁰ EP, pp. 33-4.

'reasonable' could reasonably reject is different from what people could reasonably reject if they were not reasonable. The thought is that reasonable people are motivated in a specific way, and this is partly what determines which principles and moral conceptions applicable to them will be just. So here we distinguish between the sense of reasonable in 'reasonable individuals' and the sense of reasonable in 'reasonable rejection'. The first instance of 'reasonable' would be cashed out as something like: 'committed within reason to modifying their claims, requirements, and motives in a direction which makes a common framework of justification possible'. The substance of reasonable rejection would be cashed out in terms of the objective strength of the reasons for and against principles that individuals with appropriately modified motives could give.

This would seem to permit modal contractualists to drive a wedge between modal contractualism and the straightforward view. It would license the denial that justice could be known independently of what *reasonable* people could reasonably reject. Modal contractualists could then say that just moral conceptions are just because they meet the test of reasonable rejection by individuals as opposed to that they are reasonably rejectable simply because they are just.⁶¹

Scanlon too sometimes suggests that what he calls the 'contractualist account of moral motivation' does make a difference: some reasons for rejection are made relevant by the appropriate motivation of the contracting parties, whereas others are

⁶¹ In what follows I tend to concentrate on Scanlon's account rather than Nagel's. This is for two reasons. First, Nagel explicitly takes himself to be drawing on Scanlon's view, albeit in its earliest formulation rather than that of *WWO*. Second, Nagel's account is nothing like as completely worked out as Scanlon's, so naturally discussions of the foundations and implications of what I am calling modal contractualism tend to draw on the latter rather than the former.

made irrelevant.⁶² In other words: some considerations that might otherwise be taken to be reasons in favour of certain principles are shown not to be such reasons after all by other considerations which are brought into focus only by the motivation to find principles for the general regulation of behaviour that others, similarly motivated, could not reasonably reject. As a result, the answer to the question ‘can I reasonably reject this principle?’ changes when the motivation of the person asking it changes.

This might explain why what people *take* to be the answer to the question changes when their motivation changes. But why should what it’s *reasonable* to reject change as a result of a change in motivation? Although we may grant that these other considerations are brought into focus by the motivation, if they make it reasonable to reject a principle for someone motivated to find agreement with others similarly motivated, surely they make it reasonable to reject a principle anyway—whether or not they are in focus? In his discussion of reasonableness, Scanlon claims that having the contractualist motivation “brings other reasons in its train. Given this aim, for example, it would be unreasonable to give the interests of others no weight in deciding what principles to accept. For why should they accept principles arrived at in this way?”⁶³ But why should anyone think that it’s reasonable to give the interests of others no weight in the first place?

Nagel too fails to throw much light on this question. The nearest he gets to an elaboration of the first suggestion is this:

⁶² See e.g. *WWO*, p. 157.

⁶³ *WWO*, p. 192.

there is for each person an accommodation of his partiality which is reasonable in light of the interests and partiality of others. If he is being more partial to himself than this in what he takes or insists upon, he is being unreasonable. If, on the other hand, an arrangement does not afford him the consideration which it is reasonable for him to require, in virtue of the partiality to himself which is permissible even in light of a due regard for the interests and partiality of others, then he is not unreasonable to reject it⁶⁴

The sense in which this man is being unreasonable if he is more partial to himself than a reasonable accommodation would allow might be the one we are interested in with respect to the first suggestion. That is, it seems that one is unreasonable in this sense if one doesn't modify one's 'claims, requirements, and motives'. Whereas the sense in which *others* are being unreasonable if they impose an arrangement which does not afford our man reasonable consideration may be different. It may be, for example, that the others have the appropriate motivation but simply fail to appreciate the force of our man's complaint. For our man to insist on a scheme which correctly takes into account his position would be for him not necessarily to urge that they alter their motivation—i.e. not necessarily to accuse them of unreasonableness in the first sense—but instead perhaps to appreciate that his position has not sufficiently been taken into account.

The suggestion was, recall, that what distinguishes modal contractualism from the straightforward view is the fact that the modal contractualist reasonable rejection test brings to bear a conception of appropriate motivation on the outcome of the test. What's just can't be determined without asking what appropriately

⁶⁴ EP, p. 38.

motivated people could reasonably reject. But what Nagel says about the man who is too partial to his own interests seems not to help with the question I posed a moment ago: why should anyone think it's reasonable to give the interests of others no weight in the first place? The challenge is this: if independent facts about what is best supported by reason ground the judgments about what it is reasonable to reject, then we would see no difference between what it is reasonable to reject regardless of motivation and what it is reasonable to reject given the appropriate motivation. But why would we suppose that what it is reasonable to reject regardless of motivation would not take into account appropriate consideration of, and limits on the amount that justice should cater to, the partiality of each person, both of which are supposed to be secured, on Nagel's account, only by the presence of the appropriate motivation? It seems implausible to think that what it's reasonable for me to reject even if I'm not appropriately motivated would fail to capture an appropriate balancing of individuals' partiality.

One idea that might be thought to help is that what's reasonable varies in the light of one's aims in the way that what's rational does. Now Scanlon explicitly rejects the idea that what's reasonable just is what's rational given a specified aim. Unfortunately, he doesn't say anything about whether what's reasonable *given a specified aim* is what's rational given a specified aim. At the crucial point in the discussion, he reverts to talking about what's reasonable *simpliciter*.⁶⁵ This is of no help, since we know already that Scanlon thinks what's reasonable *simpliciter* is to be motivated to find agreement on principles that no one who is similarly motivated can reasonably reject. The question is whether to be thus motivated *makes it* reasonable to give the interests of others weight in a way that it wouldn't have been

⁶⁵ See *WWO*, pp. 192-5.

without the motivation—whether others' interests provide grounds for reasonable rejection in a way that they don't if one isn't appropriately motivated. If not—if reasonableness requires one to take into account the interests of others even if one is not appropriately motivated—then the claim that no one can reasonably reject some principle given the contractualist motivation hasn't been shown to amount to more than the claim that no one can reasonably reject that principle anyway. And so the claim that we have reason to govern ourselves by principles that no one who was motivated to find agreement on principles that no one similarly motivated could reasonably reject hasn't been shown to amount to more than the claim that we have reason to govern ourselves according to principles which no one can reasonably reject *simpliciter*. Everything turns on which principles could be reasonably rejected regardless of motivation.

The idea underlying the first suggestion is that reasonable people are willing to compromise, to find an accommodation between their partial commitments and those of others. This is the sense in which they are willing to change their 'claims, requirements, and motives'. This seems right.⁶⁶ The problem with this is that as long as the kinds of facts that make it unreasonable of them not to compromise or moderate unrestrictedly partial claims are facts about how others would do under principles which satisfied those unrestricted claims, as both Scanlon and Nagel take them to be, then they seem to determine the content of the principles that it would be unreasonable of anyone to reject regardless of whether or not in fact they *are* motivated in the appropriate way. So we still haven't seen anything to license the distinction between modal contractualism and the view that what's right or just is

⁶⁶ I argue that the hybrid contractualist view that I go on to develop below captures this idea better than modal contractualism does. See Chapter 3, section 4 below.

simply that which is best supported by reason. The modal contractualist account of motivation adds nothing substantial to the content of modal contractualism. The principles that it would be unreasonable for you to reject if you were motivated to find principles which no one could reasonably reject who was similarly motivated are the same as the principles that it would be unreasonable for you to reject if you were not so motivated.

The second suggestion about the distinctiveness of modal contractualism that might be gleaned from the passage I quoted above is based on the idea that we can distinguish between the thought that just principles are those that *no one* could reasonably reject and the thought that just principles are those that none of *these* people (e.g. the people in *this* society), with their different particular perspectives, could reasonably reject. Of course if *no one* could reasonably reject some principle, then it follows that none of *these* people could reasonably reject it either. But there might also be principles that someone imaginable could reasonably reject but that none of *these* people could reasonably reject. These principles, so goes the suggestion, would attain the kind of unanimity that Nagel wants even though they might fail to attain “the kind of ideal unanimity that simply follows from their being a single right answer which everyone ought to accept because it is independently right.” Thus modal contractualism would be distinct from the straightforward view, since it might permit moral conceptions that *could* reasonably be rejected—only not by any of the people in the relevant society.

The trouble with this suggestion is that it produces a view which is simply implausible. Think about the kinds of consideration that ground reasonable rejection. Nagel notes that one basis for reasonable rejection is that a system

“leaves [the person who rejects it] too badly off by comparison with others”.⁶⁷ Scanlon suggests others, including that proposed principles “arbitrarily favor the claims of some over the identical claims of others: that is to say...they are unfair.”⁶⁸ So our second suggestion about what is distinctive to modal contractualism with respect to these kinds of considerations would be that in some circumstances none of *these* people could reasonably reject some proposed moral conception on grounds of how badly off they were or that they were arbitrarily disfavoured even though (conceivably) *someone* could reasonably reject it on those grounds.

Let us try to imagine such a case. Suppose that some proposed moral conception is such that no one under it is particularly badly off, but that were some other, much more badly off person to be a member of society, then she would be (and remain) so badly off that she would have, on Nagel’s view, grounds for reasonable rejection of the moral conception. (The content of the moral conception would not ensure, that is, that her introduction would coincide even in due course with her ceasing to be this badly off.)⁶⁹ None of the people that there are in this society, as it happens, could reasonably reject the conception, but if we change things a little, at least one of them could—so it’s not the case that *no one* (conceivable) could reasonably reject it.

Obviously we must abstract away from questions about *why* the badly off person would be badly off to the following degree: we must assume that she would not be badly off for reasons which we view as undermining the grounds for reasonable

⁶⁷ EP, p. 38.

⁶⁸ WWO, p. 216.

⁶⁹ So we should probably not think of her introduction as being effected through immigration, for example, since there may be particular principles which govern immigration and undermine immigrants’ claims to the same treatment as others. See the following paragraph in the main text.

rejection that being badly off provides. So she would be badly off in a way which we, as modal contractualists, would think a moral conception ought not to permit. This means that the fact that no member of this society is *as it happens* sufficiently badly off to reject it is a contingent matter. It is not thanks to the content of the moral conception itself that no one is in a position to reject it, since the moral conception does not include provisions to prevent its coming about that anyone should be this badly off.

An example of the sort of case in question (which also illustrates the point with respect to unfairness as a grounds for reasonable rejection) would be as follows. An ethnically completely homogeneous society has a constitution which is racist, the legacy of older times of ethnic diversity and divisions. As it happens, no one is affected by the constitutional racism, since no one belongs to the races it discriminates against. But if someone from the discriminated-against races were introduced into society, she would be far worse off than everyone else thanks to provisions against people from those races taking jobs, receiving benefits, and so on.⁷⁰ Now, it strikes me as implausible to suppose that modal contractualists would want to endorse moral conceptions like this, even for societies in which no one as a matter of fact falls foul of the racist constitution ~~or~~ has a disability. So if that is what the suggestion under review involves, then it should be rejected. And how could it avoid involving that?

Someone might object that I am illicitly supposing here that grounds for reasonable rejection are restricted to only those grounds that individuals in the relevant group have themselves or can envisage themselves, the actual members of the group that they are, having. But they are not to be thought of as restricted in

⁷⁰ This example illustrates the point with respect to unfairness as well.

this way: in considering what could be reasonably rejected, individuals should “rely...on commonly available information about what people have reason to want” rather than on what *they alone* want or are likely to want.⁷¹ People from different ethnic backgrounds and people with disabilities are people, and so the reasons that they would have to reject the moral conceptions in question should also be taken into account even by members of societies in which no one differs in terms of ethnicity or ability.

But if this is right then we lose sight of the sense in which modal contractualism is supposed to be distinctive on the second suggestion. What none of *these* people could reasonably reject is surely now the same as what *no one* could reasonably reject, and what cannot be reasonably rejected on the modal contractualist view has not, therefore, been shown to be different from what is best supported by reason. For considerations from whose standpoint are now excluded by the modal contractualist formula? Anything that *someone* could reasonably reject is something that *one of us* could reasonably reject, since we are required to rely on the reasons that everyone has rather than those that *only we* have. (Of course there is an analogue of the actual contractualist variable of scope to consider here: who counts as someone? But *anyone* who counts as someone will be such that reliance on generic reasons in working out what someone could reasonably reject will automatically take her perspective into account.) I conclude, then, that the second suggestion does not vindicate the thought that modal contractualism is distinct from the view that what is just is that which is best supported by reason.

The final suggestion that I want to consider about the distinctiveness of modal contractualism is as follows. Scanlon says that his contractualism accounts for the

⁷¹ *WWO*, p. 204. Scanlon calls such considerations ‘generic reasons’.

special significance of the range of conclusions that it supports because it “recognise[s] that what is at stake are the reasons we can offer one another in a process of mutual justification.”⁷² It’s not obvious, however, what difference recognising this is supposed to make, or indeed that justifying principles by appeal simply to reasons that we take there to be, as on the straightforward view, fails to recognise it. One thing that Scanlon says is that his contractualism accounts for the special significance of the moral because it explains the sense in which failure to recognise moral reasons is so much more important than a failure to recognise other reasons that there are, viz., because it reflects a person’s “failure to see why the justifiability of his or her actions to us should be of any importance”.⁷³ But if it’s true that what’s unjustifiable on the straightforward view is unjustifiable to each person, then it’s not true that the principles picked out by modal contractualism are only moral ones so long as what’s unjustifiable on the straightforward view includes some actions which aren’t immoral. In that case, modal contractualism doesn’t account for the special significance of the *moral*. If, however, what’s unjustifiable on the straightforward view includes only immoral actions, then contractualism does account for the special significance of the moral. But then so does the straightforward view. It says (something like) that failure to be moral reflects failure to care about being justified in one’s actions.

Scanlon’s suggestion can, however, be given more substance by the following idea. Scanlon has claimed that it is essential to contractualism not only that it expresses the idea of justifiability to each person but also that it respects what Parfit

⁷² Scanlon, “Replies”, p. 434.

⁷³ *WWO*, p. 159. Cf. Stephen Darwall, “Contractualism, Root and Branch: A Review Essay”, pp. 204-207.

has called the Individualist Restriction.⁷⁴ According to the Individualist Restriction, “[i]n rejecting some moral principle we must appeal to this principle’s implications only for ourselves, or for any other *single* person.”⁷⁵ Modal contractualist theories which respect the Individualist Restriction hold that a principle satisfies the ideal of individualistic justification only if its implications for any given single person are not such that it could be rejected just on the grounds of those implications. For any principle justifiable to each person, it follows that the reasons for it are stronger than the reasons against it, where the range of reasons for or against is restricted to implications for single persons.⁷⁶

⁷⁴ It’s not entirely clear whether or not he continues to claim this. See “Replies” pp. 432-4. But without it there seems to be nothing distinctive about modal contractualism at all. I know of no other suggestion which might explain how it differs from the straightforward view.

⁷⁵ Parfit, “Justifiability to Each Person”, p. 372. See also *WWO*, p. 229: “a central feature of contractualism [is] its insistence that the justifiability of a moral principle depends only on various *individuals’* reasons for objecting to that principle and alternatives to it.”

⁷⁶ One way to think about the restriction to implications for a single person would be as a restriction to implications in terms of how badly off the person is (and how badly off she would be under alternative principles and how badly off she is compared to everyone else). This is the ‘complaint model’ of (modal) contractualism, which Scanlon appeared to endorse originally (see “Contractualism and Utilitarianism”, pp. 144-5) but has since explicitly rejected. (Nagel, who argues that contractualism supports an approach under which we pay special attention to the worst off under proposed principles, can be interpreted as endorsing the complaint model. See *EP*, pp. 63-74.) According to Scanlon’s more recent formulations, considerations of unfairness, responsibility, and other moral notions can involve implications for single individuals which may thus be invoked as grounds for reasonable rejection without violating the Individualist Restriction. (I say that these notions involve ‘implications for single individuals’ rather than following Scanlon in saying merely that they are ‘individuals’ reasons’ because I find the latter phrasing obscure.) See *WWO*, pp. 213-8 and 229.

The Individualist Restriction makes modal contractualism less confusing.

Because the restriction rules out some kinds of reasons which might justify some principles (i.e. those reasons not based on the implications for single persons), some views which might be justified if such reasons were permitted cannot be justified on an Individualist Restriction-respecting modal contractualist account. To go back to our earlier argument, for example: one reason that someone might give in support of some form of utilitarianism is that more utility will be created thereby. But even if this is true, and even if we concede that utility is the only good, utilitarianism does not meet the ideal of individualistic justification according to an Individualist Restriction-respecting modal contractualism unless its being or not being implemented counts non-circularly as an implication for single persons.⁷⁷ Now, as we saw, one notorious consequence of utilitarianism is that one person's life may be sacrificed if enough other people would benefit in some small way as a result. The strongest reason to reject a utilitarian principle that its implications for any single person give rise to might well be, therefore, that of the sacrificed person, who would lose her life. The strongest reason to reject a non-utilitarian principle (as compared only with a utilitarian principle) that its implications for any single person give rise to might well be, meanwhile, that of one of the people who would benefit in a small way if the utilitarian principle were implemented. So the strongest reason that any single person has for rejecting a non-utilitarian principle is much weaker than the strongest reason anyone has to reject a utilitarian principle. Therefore a utilitarian moral conception is not acceptable according to an Individualist

⁷⁷ The non-circularity stipulation means that the failure to implement utilitarian principles couldn't count as an implication for single individuals just because it would be *wrong* or *unjust* (that is, just because it would do them a wrong or an injustice). For that would be to presuppose precisely what we are trying to give an account of.

Restriction-respecting modal contractualism.⁷⁸ Here, then, is a way in which modal contractualism can be clearly distinguished from the straightforward view. Modal contractualism thus distinguished is a substantial view in that it holds certain considerations that we might otherwise suppose to be reasons in favour of a principle or moral conception to be weightless.⁷⁹

Whether or not any given contractualist actually accepts the Individualist Restriction, it can be seen that it's a fairly natural restriction for modal contractualists to impose. If you distinguish between individualistic justification and justification *simpliciter*, taking the former to be fundamental in the case of justice, it's natural to suppose that what you want to know is whether that principle's implications for any given single individual should be accepted, given the implications for other single individuals of alternative principles. It has been suggested by some philosophers that the Individualistic Restriction is inessential to what I am calling modal contractualism. Parfit, for example, claims that it is plausible that "in rejecting some principle, each person could appeal to the burdens that this principle would impose not only on her, but also to other people" and concludes that "Scanlon could give up his Individualist Restriction without giving up, or in any way weakening, his idea of justifiability to each person."⁸⁰ David Brink

⁷⁸ See Scanlon, "Contractualism and Utilitarianism".

⁷⁹ Aggregative considerations are those most clearly ruled out by the Individualist Restriction, but there may be others (Scanlon rules out impersonal considerations, for example, at *WWO*, p. 200). Nagel identifies himself with Scanlon's contractualism but nevertheless appears to accept some aggregation (see "Equality", p. 125, and *EP*, p. 68). However, he seems uneasy about the compatibility of this with the pairwise comparison that he takes to be fundamental to the idealised unanimity realised by (what I am calling) modal contractualism.

⁸⁰ Parfit, "Justifiability to Each Person", p. 388.

argues that allowing aggregative considerations as the basis for reasonable rejection comes closer to the ideal of unanimity than does an Individualist Restriction-respecting modal contractualism.⁸¹ But the putatively modal contractualist view that Parfit suggests instead appears to have nothing to distinguish it from the straightforward view. It appeals simply to the truth of distributive principles such as the Priority View (see above). Brink, meanwhile, describes a view which he himself takes to be non-contractualist which is, effectively, Scanlonian contractualism without the Individualist Restriction. So given my analysis of Scanlon and Nagel, the most prominent contemporary modal contractualists, and despite these philosophers' claims to the contrary, I shall assume from now on that the distinctive feature of modal contractualism is its respect for the Individualist Restriction.⁸²

4. *First-order normative modal contractualism*

What would be involved in building a first-order normative modal contractualist theory? I have claimed that what is distinctive about modal contractualism, what distinguishes it from the view that the just moral conception is that which there is least reason to reject, is its requirement that reasons for the rejection of a moral conception respect the Individualist Restriction. This means that certain kinds of arguments for moral conceptions are ruled out, and that in turn means that reasons to suppose that those moral conceptions should be preferred to other moral

⁸¹ See Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory", pp. 269-70 and 279-80.

⁸² See Michael Otsuka's "Saving Lives, Moral Theory, and the Claims of Individuals" for further arguments supporting this assumption.

conceptions may turn out to be inadmissible in a modal contractualist first-order theory.

But once the Individualist Restriction is taken into account, formulating a first-order modal contractualist theory is simply a matter of identifying and comparing the (Individualist Restriction-respecting) reasons that there are in favour of competing moral conceptions. For illustration, consider the following three possible moral conceptions: utilitarianism, strict egalitarianism, and what I shall call 'arbitrary prioritarianism'. Utilitarianism is familiar enough in broad outline to need no further explanation for the purposes of this example. Strict egalitarianism requires that each individual have exactly the same amount of each relevant good. Let us stipulate that as it turns out this is very little, perhaps because Rawls-style incentives are necessary to increase even the smallest share, or because 'lumpy' goods are simply thrown away. Arbitrary prioritarianism works as follows: the share of the worst off is maximised as a result of a Rawls-style difference principle, but only once a third of the population is arbitrarily excluded (on the basis of earlobe length, say) from taking up the incentives which are used in order to increase the total sum of goods and thus the minimum share. Thus, because everyone is more or less equally talented, the worst off under arbitrary prioritarianism are better off than under either strict egalitarianism or utilitarianism.

Now, the chief consideration in support of utilitarianism over egalitarianism and arbitrary prioritarianism appeals to its maximisation of utility, even though that may involve the sacrifice of some people for the sake of a small increase in utility per person remaining. Arbitrary prioritarianism is chiefly supported by the consideration that the worst off would do a lot worse under the other two principles, but clearly it is arbitrarily discriminatory. (We focus on the worst off because she is likely to be the individual with the strongest reason for rejection

under any given conception.)⁸³ Strict egalitarianism is not (let us say) unfair, but evidently it fares worse than arbitrary prioritarianism in terms of the level of well-being of the worst off. Assume that there are no other relevant dimensions of assessment besides these three and that no other moral conceptions may be proposed.

As we saw in the preceding section, the chief consideration in support of utilitarianism violates the Individualist restriction. No *single* individual can claim that her loss (of the small increase in utility gained as a result of the sacrifice of one person under utilitarianism) under arbitrary prioritarianism or strict egalitarianism provides her with a reason to reject either of those moral conceptions which is stronger than the sacrificed person's reason to reject utilitarianism. The fact that the *aggregate* loss under the competing conceptions may be greater than the sacrificed person's loss is ruled out as irrelevant by the Individualist Restriction. (The point is *not* that it is outweighed.) So insofar as this is the reason that utilitarians can give in support of their view against arbitrary prioritarianism and strict egalitarianism, it has *nothing* to say in its favour for a modal contractualist. But as regards *fairness*, utilitarianism does do better than arbitrary prioritarianism in a dimension which respects the Individualist Restriction:⁸⁴ it does not arbitrarily discriminate against anyone in its calculus. So in that dimension there may be a tie between utilitarianism and strict egalitarianism. But in the choice between those two alone strict egalitarianism must win, since the only reason that utilitarians can give against strict egalitarianism is ruled out altogether.

⁸³ See Scanlon, "Contractualism and Utilitarianism", p. 145.

⁸⁴ At least according to the complaint-model rejecting Scanlonian view.

As regards the choice between strict egalitarianism and arbitrary prioritarianism, however, it's not clear what a modal contractualist should say. Presumably the worst off individual under strict egalitarianism has a reason for the rejection of that conception, based on how well she fares, which is stronger than that of the worst off individual under arbitrary prioritarianism based on how well *she* fares. But on the other hand, no one under strict egalitarianism is arbitrarily discriminated against, so those who are arbitrarily discriminated against under arbitrary prioritarianism will have a reason for its rejection based on that consideration which finds no counterpart under strict egalitarianism.

Formulating a modal contractualist first-order normative theory given only these three conceptions and the three dimensions of assessment to choose from is, then, a matter first of determining which dimensions of assessment (reasons for moral conceptions) do not violate the Individualist Restriction and then determining, on the basis of those that remain, which theory is best supported. In our example this involves ruling out the chief reason in support of utilitarianism and then determining which of the three conceptions is best supported by reasons of fairness and the well-being of the worst off. Utilitarianism is poorly supported by either reason in comparison with the other two, so it will be ruled out. Perhaps then we will decide that having the level of well-being that the worst off person under strict egalitarianism has gives one a reason for the rejection of that moral conception which is stronger than the reason that being arbitrarily discriminated against gives the worst off short-earlobed individual under arbitrary prioritarianism to reject *that* moral conception. Perhaps we will not. But what is clear is that this will be a matter simply of discerning and comparing the relative strength of the (Individualist Restriction-respecting) reasons that there are.

The same holds outwith the artificial constraints of this example. The range of moral conceptions available to choose from is much larger, of course, and there are other Individualist Restriction-respecting reasons for rejection which cannot be captured in terms of individual levels of well-being or arbitrary discrimination. Scanlon suggests, for example, that grounds for rejection may be given by considerations of rights and entitlements, of individual responsibility, and of urgency of need.⁸⁵ But the same basic approach applies: consider the Individualist Restriction-respecting reasons that may be offered for and against moral conceptions, and choose that conception which is least rejectable, in the sense that the strongest reason against it is weaker than the strongest reason against any competing conception. So, for instance, when Scanlon proposed in “Contractualism and Utilitarianism” that Rawls be read as advancing what I am calling a modal contractualist theory, he based his reading on the idea that certain conditions on the parties’ choice in Rawls’s original position are best viewed as expressing “the strength of the objection that the ‘losers’ might have to a scheme that maximised average utility at their expense, as compared with the counter-objections that others might have to a more egalitarian arrangement.”⁸⁶ The thought underlying the construction of the original position is simply that once we have discounted Individualist Restriction-violating reasons, the strongest reason against average utilitarianism here is weaker than the strongest reason against ‘a more egalitarian arrangement’. The original position is to be interpreted as a means of ensuring that our choice is guided by the Individualist Restriction-respecting reasons that there are.

⁸⁵ See *WWO*, pp. 214, 223-8, and 248-93.

⁸⁶ “Contractualism and Utilitarianism”, p. 148.

So what will a modal contractualist theory of justice look like? That will depend upon which Individualist Restriction-respecting reasons we take there to be. Reasons relating to levels of individual well-being will surely be among them, as they are in Scanlon's and Nagel's views. But whichever such reasons we do take there to be, a moral conception's conformity to them will be necessary and sufficient for justice. If the strongest reason against some egalitarian moral conception is weaker than the strongest reason against some competing moral conception, then to the extent that our choice is restricted to those two moral conceptions, the egalitarian moral conception is the more just.

5. Modal contractualism and actual contractualism

Could a modal contractualist be at the same time an actual contractualist? That depends on whether the consideration appealed to by actual contractualists as grounds for the rejection of proposed moral conceptions produces reasons which do not violate the Individualist Restriction. The consideration is simply that a relevant individual does not actually accept the moral conception in question. As a reason for the rejection of that conception, this does not violate the Individualist Restriction. An appeal to my actual non-acceptance of a moral conception may not be an appeal to facts about my well-being. But it can be seen as analogous to an appeal to a moral conception's arbitrary treatment of me in the sense that it is an appeal to the principle's implications for *me*, a single individual.⁸⁷ It is not an appeal

⁸⁷ Even if we choose to endorse the complaint model, so that arbitrary treatment of me is *not* clearly a reason for rejection, still my actual non-acceptance might be thought to provide grounds for

to the implications of the moral conception for any aggregate group. So a modal contractualist could identify non-acceptance as a reason for rejection without losing her grip on what is distinctive about modal contractualism.

Of course not many modal contractualists *would* identify non-acceptance as a reason for the rejection of a moral conception—much less the only one, as a pure actual contractualism founded on modal contractualism might. But it would not obviously be inconsistent with modal contractualism for them to do so. One difficulty, however, is that if there is no moral conception that each person actually accepts, the actual contractualist claims that no moral conception is just. That would translate into the conclusion that each could reasonably be rejected. But Scanlon seems to deny that such cases, “in which opposing parties have strong objections that are evenly balanced”, could in fact arise.⁸⁸ If modal contractualism implies Scanlon’s denial, modal contractualism and actual contractualism are incompatible.

Scanlon’s argument against the possibility of evenly balanced objections is not particularly powerful, however. He imagines two people swimming from a sinking ship, one of whom is much stronger than the other, who chance upon a single lifejacket at the same time. He then contemplates the idea that both a principle permitting the use of force to seize the lifejacket and a principle prohibiting the use of force could be reasonably rejected (one swimmer’s objection to one principle is as strong as the other’s to the other), since “the considerations on the two sides are

reasonable rejection because living my life in a world structured by a normative system that I do not accept has severe implications for my well-being.

⁸⁸ *WWO*, p. 196.

the same”.⁸⁹ But he rejects this idea on the ground that in fact one of these principles, perhaps combined with others, will best recognise the “need for some decisive solution” as well as the symmetry of the swimmers’ claims. More generally, he says, “[t]he very fact that...objections are symmetrical may point the way toward a class of principles that are not rejectable.”⁹⁰

To start with, however, that is consistent with the possibility that the fact that objections are symmetrical may *not* point the way towards such a class of principles. In the swimmers’ case, indeed, the symmetry of their objections and urgency of their needs might point a way to non-rejectable principles, as Scanlon claims. Yet perhaps not all analogous situations point to analogous non-rejectable principles. A situation in which each possible moral conception is actually rejected by *someone* may be like this. (And that might simply be a possibility we must live with.)

In any case, it’s not clear what Scanlon means when he says that “[a] principle permitting each to struggle for the jacket at least has the merit of recognizing the symmetry of their claims”.⁹¹ Why wouldn’t a principle prohibiting them from struggling have just as much merit in that sense? Why wouldn’t a refusal to endorse any norm pertaining to this situation have just as much merit? Scanlon’s more significant point seems to be that a principle permitting each to struggle for the jacket has the merit of “recognizing...the need for some decisive solution”. Here it seems that a principle prohibiting the struggle is weaker.⁹² But it’s not clear at all

⁸⁹ *WWO*, p. 196.

⁹⁰ *WWO*, p. 197.

⁹¹ *WWO*, p. 196.

⁹² Although this may overestimate the decisiveness of violent struggle in determining a ‘winner’ (and underestimate the decisiveness of, say, calmly treading water and hoping that the other person will tire first).

that in this situation there *is* a need for a decisive solution in the sense of an outcome which receives morality's seal of approval (by being the outcome of a non-rejectable principle). Perhaps rightness and wrongness (or justice and injustice) don't extend to every act in every situation. If the parties go ahead and struggle with each other we might want to say afterwards that they did what they had to do, but we might not want to say all the same that the winner did no wrong *or* that there was some solution which would have been right. So although a hybrid of modal and actual contractualism raises the possibility of a situation in which every possible moral conception could be rejected for the same reason—actual non-acceptance by some individual—I conclude that this does not necessarily threaten such a view. Modal contractualists needn't follow Scanlon in denying the existence of such situations.

Is modal contractualism consistent with *all* of the forms of actual contractualism that I identified in chapter 1? In particular, mixed and dual condition theories may appear to be problematic. According to these versions of actual contractualism, it is a condition of the justice of a moral conception that at least some (all, in dual condition actual contractualism) parts of that conception both are actually accepted by each person in society—the legitimacy component—and meet conditions independent of such acceptance that, we suppose, those parts of a just moral conception ought to meet—the regulative conception component. Could a modal contractualist consistently incorporate both of these conditions into her account of the justice of a moral conception? There are a number of ways in which she might try to do so, but we need only describe one in order to see that it could be done.

One way, then, in which someone might incorporate both conditions would be to account for both the legitimacy component and the regulative conception component in a modal contractualist framework. This would involve viewing actual

non-acceptance as a basis for reasonable rejection. At the same time, failure to be in accordance with the moral conception set out in the regulative conception component would also be viewed as a basis for reasonable rejection. We would say of the regulative conception component that it meets the modal contractualist test of reasonable rejectability notwithstanding the reasons provided by actual non-acceptance. The reasons provided by anyone's actual non-acceptance and the reasons provided by the independent considerations in favour of the regulative conception would be said to be evenly balanced.⁹³ The result would be that only the

⁹³ It might be argued that the reasons invoked by the regulative conception component must add force to the complaint against other moral conceptions which is constituted by a person's non-acceptance of it, so that complaints against the moral conception endorsed in the regulative conception component will always be outweighed by complaints against other moral conceptions. In that case, the possibility arises that a person's non-acceptance of some other moral conception might be insufficient for its rejection. If this is right, incorporating both actual contractualism and the regulative conception component into the modal contractualist framework may involve treating the reasons invoked by each of these components as 'threshold' reasons, each sufficient, on its own, that is, to ground the rejection of a moral conception.

Scanlon denies the possibility of such reasons at *WWO*, p. 195. The basis of his denial seems to be the claim that there are no cases in which every principle can be reasonably rejected, because considerations apart from considerations of well-being will act as tie-breakers. But this sounds like an argument against the possibility of evenly balanced objections, not against the threshold view. So perhaps Scanlon's argument is this:

1. Even in cases where a person stands to lose her life whatever principle we adopt there will be considerations which make it reasonable to favour one principle over the alternatives
2. The loss of life is not always sufficient for reasonable rejection
3. If the loss of a person's life is not always sufficient for reasonable rejection then nothing can be
4. The threshold view is false

moral conception selected by the regulative conception component could meet the reasonable rejection test. But even that could not do so unless each person actually accepted it.

So far I have argued that there is no reason to suppose that actual contractualism and modal contractualism are incompatible, even when actual contractualism takes one of the more complicated forms that I identified in Chapter 1. And someone might indeed hope to make a hybrid view plausible. For she might both accept the Individualist Restriction and want to contend that as a matter of justice someone's actual non-acceptance of a moral conception should always be considered reasonable grounds for the rejection of that conception. What I want to do now is turn to the more specific question of what such a hybrid view would say. I shall argue in particular that hybrid theories face two puzzles which they must solve in order to be plausible.

6. Two puzzles for hybrid theorists

Suppose, then, that we are moved both by the modal contractualist's interpretation of the contractualist ideal of individualistic justification and by the idea that individual actual non-acceptance is sufficient to show that a moral conception is not

If this is not Scanlon's argument, I am not sure what is. He says nothing else about the threshold view. But although this is not the argument against the possibility of evenly balanced objections, it clearly depends on that argument. I rejected that argument above, so I conclude that Scanlon's denial does not show that it is fatal to hybrid contractualism if it must treat as threshold reasons the reasons it identifies.

just. We therefore develop a hybrid theory according to which individual actual non-acceptance is grounds (and for simplicity let us say the only grounds)⁹⁴ for reasonable rejection of a moral conception. We say: it is a (sufficient) Individualist Restriction-respecting reason to reject principle or moral conception P that individual X actually does not accept P.

So far so good. But now the following complication arises. Modal contractualists suppose that everyone has reason to find and comply with—to view, ordinarily, as overriding—principles that cannot reasonably be rejected.⁹⁵ Being motivated by this reason is part of being reasonable.⁹⁶ This motivation may make a difference to whether or not someone actually accepts proposed principles and moral conceptions. Now, in an ordinary modal contractualist schema this is of no consequence. Utilitarianism can reasonably be rejected, Scanlon (for example) might say, since the fate of the worst off individual under utilitarianism is far worse than the fate of the worst off under some favoured alternative moral conception. The fate of the worst off gives grounds for reasonable rejection which are clearly unaffected by whether or not she herself or anyone else happens to be appropriately motivated. So there is no question that utilitarianism can be reasonably rejected on these grounds.

But in a hybrid theory things are not so clear. As we saw a moment ago, a hybrid theorist thinks that it is a (sufficient) reason to reject principle or moral

⁹⁴ So effectively we are considering the hybrid theory as a pure legitimacy view for the time being.

⁹⁵ See above. I claimed that the specification of the motivation to comply with this reason in the modal contractualist formula is unnecessary, but that was because I take it that it makes no difference to the principles picked out by that formula, not because modal contractualists should not suppose that we do indeed have this reason.

⁹⁶ See *WWO*, p. 157.

conception P that individual X actually does not accept P. But suppose that X is, as it happens, unreasonable in the sense that she is not motivated to comply with the reason that everyone has to find and comply with principles that cannot be reasonably rejected. That means, for a hybrid theorist, that she is not motivated to find and comply with principles that everyone actually accepts. But whether or not she has this motivation might make a difference to whether or not she actually accepts P. It might be that if she *had* been appropriately motivated she would have actually *accepted* P.

Why might we suppose that X's compliance with that reason might entail such acceptance? Well, suppose that you give the reason that you have to find and comply with principles that cannot be reasonably rejected priority in your practical reasoning. To repeat: this means, on a hybrid theory, that you give priority to finding and complying with principles which, among other things, everyone actually accepts. Now had you not given this reason priority in this way, there might have been principles that everyone else actually accepted but you did not, because they conflicted with your own moral outlook. But since you do give it priority, the conflict with these other norms is less important to you than finding and complying with principles that everyone actually accepts. Since compliance with the reason that you have to do this will be made possible by your actual acceptance of the principles which otherwise you would not have accepted, you might well take yourself to have reason actually to accept these principles after all, and you might therefore actually accept them.⁹⁷ The more weight you give to the reason that you

⁹⁷ See Scanlon's discussion of the way that acceptance of this basic reason discounts as any kind of reason at all certain considerations which otherwise would have counted as reasons at *WWO*, pp.

have to find and comply with principles that cannot reasonably be rejected, the more likely such an outcome is. If you give it absolute or near-absolute weight—as indeed we might suppose a modal contractualist would think you should, since this is a matter of justice or what we owe to each other—then it is almost certain. You will almost certainly be ready to discount any normative conflict that proposed principles provoke for you in order to find agreement on principles which everyone actually accepts. And the same goes for everyone else.

If this is right, I think that the hybrid theorist gives substance to one claimed feature of modal contractualism but also faces a puzzle. The claimed feature is that it makes a difference to the content of what it is reasonable to reject that we are interested in those who are appropriately motivated—those who *take* themselves to have reason to find and comply with principles that no one could reasonably reject. This is an attractive idea in itself: that justice proceeds from an appropriate reconciliation of the interests of *reasonable* people, those who are ready to adjust their claims in the light of the reasonable claims of others, rather than proceeding from a balancing of the interests of everyone even when they are unreasonable. But a hybrid theory gives this idea some substance. Above I argued that as long as the facts that determine whether or not it is reasonable for someone to reject a principle are facts about how people would do under it and competing principles, then whether or not any given individual is motivated to find and comply with principles that no one could reasonably reject makes no difference to the content of such principles. Whether she finds them and complies with them or not, the reasons are there, independent of her motivation. So there is no distinction between justice as a reconciliation between reasonable people and justice as a balancing of all individuals' interests, reasonable or not. But now we see that being appropriately motivated *can* make a difference: when the facts that determine the reasonableness

of rejecting a principle are facts about how that principle fits with one's overall moral views, including one's commitment to finding and complying with principles that no one can reasonably reject. So a hybrid view can make good, along with the Individualist Restriction, the modal contractualist's claim that modal contractualism is more than a view that says what is just is simply that which is best supported by the reasons. What the reasons *are* is not independent of individuals' reasonableness. Finding and complying with the reasons that there are requires that we actually do aim to find and comply with such reasons.

Now the puzzle. At least insofar as we are considering only individuals who are reasonable in the sense that they give priority in their practical reasoning to the reason that they have to find and comply with principles that everyone actually accepts, it appears that what is just has nothing to do with anyone's moral (or other) views and commitments aside from the commitment to the modal contractualist reason itself.⁹⁸ For—so long as it's given sufficient priority—the reason that an individual has to seek and comply with universally accepted principles will trump any reluctance based on her *other* views and commitments to comply with principles that look likely to gain universal acceptance. This will be true of each individual. No one's views and commitments apart from the commitment to the modal contractualist reason can be taken to be fixed, a basis for the acceptance or rejection of principles. Meanwhile, the modal contractualist reason in itself doesn't determine any first-order principles. So the content of just principles and moral conceptions may be determined by factors which would strike us as arbitrary and so inappropriate. The reasons identified by the modal contractualist component of the

⁹⁸ Recall the stipulation above that we are considering the hybrid theory as a pure legitimacy theory for the time being.

hybrid theory do not determine their content. Nor do the values and principles affirmed by individuals—which is where we would expect to find the determining factors in an actual contractualist view. Yet a self-appointed co-ordinator's first moral conception out of a hat, the arbitrary choice of a randomly nominated person, the principle that the majority would accept were considerations of reasonableness not taken into account—all of these might determine the content of justice. Surely this counts against the hybrid contractualist?

Of course, if we drop the stipulation that we are considering hybrid theories as pure legitimacy theories and bring in a regulative conception component, then it's no longer the case that any moral conception could be just. Clearly the moral conception specified in the regulative conception component is the only just choice. Anything else violates the necessary condition set by the regulative conception component. But this doesn't alter the apparent arbitrariness of the contribution from the hybrid theorist's specification of the legitimacy component. Justice could easily be secured for the moral conception specified by the regulative conception component: all that's needed is to co-ordinate individuals so that this is what they accept, rather than any other moral conception—any of which they would equally gladly accept in order to facilitate the universality of actual acceptance. But mere co-ordination isn't an appropriate determinant of the content of what's just, any more than the arbitrary choice of a randomly nominated person is. So the puzzle, which I shall call the 'arbitrariness puzzle', remains.

I said that this puzzle arises at least insofar as we are considering only reasonable people, that is, people who give priority to the modal contractualist reason. A different puzzle ('the puzzle of scope reconciliation') comes to light when we consider what the hybrid theorist should say about people who are not reasonable in this sense. Such people would not stand ready to abandon their

commitments and accept others in order to facilitate universal acceptance. Now, the hybrid theorist claims that it is a (sufficient) reason to reject principle or moral conception P that a given individual S actually does not accept P . But what if S 's actual non-acceptance of P comes about only because S fails to recognise the reason that there is to reject a principle or moral conception that someone's actual non-acceptance provides? In that case even if everyone apart from S accepts P , P —so it seems—is reasonably rejectable. But this is only because some people don't care about what can and cannot reasonably be rejected. This appears to contaminate the content of justice (what cannot be reasonably rejected) by allowing it to be determined by the views of those who are unjust (who don't care about what can or cannot be reasonably rejected). Surely it should not be so contaminated?

This brings out a question about scope. Modal contractualists and therefore hybrid theorists include among those whose *reasonable* rejection matters even those who are, as a matter of fact, unreasonable: unreasonableness is not a criterion for exclusion by the modal contractualist's counterpart to the actual contractualist

variable of scope (see Chapter 1, section 3 above).⁹⁹ They care about what unreasonable people could *reasonably* be expected to accept. But whose *actual*

⁹⁹ Scanlon discusses this question at *WWO*, pp. 177–87. In actual contractualism the question of scope is, effectively: to whom should we give an actual veto (to be used as she pleases) over proposed moral conceptions? It might be plausible to say: only to reasonable people—because, say, unreasonable people's unreasonableness should not be permitted to determine the content of justice. (See below in the main text.) In modal contractualism the question is: implications of proposed principles for *whom* should be considered among the Individualist Restriction-respecting grounds for reasonable rejection? It is not as plausible to suppose that we ought not to consider unreasonable people among those the implications for whom should be considered, since their unreasonableness does not as clearly determine the content of justice as it does for actual contractualists.

rejection matters to a hybrid theorist? If it is not only those who are reasonable, then we face the difficulty just described. But is there any basis internal to the hybrid view for excluding those who are unreasonable from the set of those individuals whose *actual* acceptance matters?

Perhaps there is. The modal contractualist ordinarily doesn't exclude those who are unreasonable from the scope of those the implications for whom of a proposed principle or moral conception may count among the grounds for the reasonable rejection of that principle or moral conception. Each person will be subject to the principles or moral conception which is ultimately chosen, and it will have implications for each person which are surely relevant to the decision what it is reasonable to reject on the basis of reasons deriving from implications for individuals. That a person is unreasonable doesn't exclude her from moral consideration. However, it does give us ground to ignore her actual judgments about what is right and wrong, just and unjust. These ought not to have a bearing on the question what is right and wrong, just and unjust. For they are, *ex hypothesi*, not based on recognition of a fundamental reason (that we have to find and comply with principles that cannot be reasonably rejected) that there is pertaining to the matter of what is right and wrong and just and unjust. What we want to know is what such people would accept if they recognised this reason, not what they do accept when they don't recognise it. Acceptance of the existence of this reason is a fundamental modal contractualist commitment.

Now an unreasonable individual's actual non-acceptance may, as we saw above, be shaped by her unreasonableness in the sense that it is only because she is unreasonable that she does not accept the principle or moral conception in question. She fails to order her norms and values in the way that reason directs. If this is so, to take her actual non-acceptance as a basis for reasonable rejection is

effectively to allow her actual judgments about what is right and wrong, just and unjust, to have a bearing on the question of what really *is* right and wrong and just and unjust. But as we just saw, modal contractualists need not do this. And since hybrid theorists are modal contractualists, since they adhere to the fundamental modal contractualist commitment—the view that we have reason to find and comply with principles which cannot be reasonably rejected—it follows that nor need they take unreasonable individuals' actual non-acceptance, insofar as it *is* informed by their unreasonableness, as a basis for reasonable rejection. The variable of scope for actual contractualists may thus be determined by their modal contractualist commitments (as well as any other plausible factors, such as minimal rationality conditions and the like). So the hybrid theorist can legitimately exclude unreasonable people from the scope of those whose actual acceptance and rejection counts as a basis for *reasonable* acceptance and rejection.¹⁰⁰

This raises the possibility, however, that the hybrid view—though consistent and evidently incorporating elements of actual contractualism—fails to respond appropriately to the fundamental actual contractualist motivation. I argued in Chapter 1 that actual contractualist restrictions of scope are unobjectionable when they are driven by fundamental actual contractualist commitments. Here, however, the restriction in scope is driven by *modal* contractualist commitments. I also said in Chapter 1 that some specifications of scope would show that the putatively actual contractualist theory in question was not, after all, actual contractualist at all. A

¹⁰⁰ What if the unreasonable person's rejection of some principle is not a product of her unreasonableness? Well, so long as we think about what compliance with the modal contractualist reason entails in the way I described above, someone's rejection of a principle regardless of what others do is always unreasonable, since a reasonable person is, as we saw, prepared to accept anything if others also accept it.

specification of scope and acceptance which restrict what we care about to what reasonable individuals do accept and what unreasonable individuals *would* accept insofar as they were reasonable sound suspicious in exactly this way. Consistent, then, it may be; but it is possible that this combination of actual and modal contractualism in fact loses sight of much of what is appealing about actual contractualism in the first place. This is the puzzle of scope reconciliation.

I am not going to attempt to solve either of these two puzzles until later. (I shall present solutions to them as part of my elaboration of the Rawlsian hybrid contractualism that I shall be endorsing.) For now I hope only to have made it clear that hybrid theorists cannot afford to ignore the obstacles they pose to a complete elaboration of the hybrid view.

7. *Conclusion*

In the first two chapters of this thesis I have presented two different contractualist frameworks which interpret the contractualist ideal of individualistic justification in different ways. Actual contractualists suppose that the ideal is satisfied with respect to a given moral conception (or aspects of one) only when each relevant individual actually does accept that moral conception. Actual contractualists can be more or less thoroughgoing in their actual contractualism in the sense that they may suppose that more or less of the moral conception implemented in our society must satisfy the ideal in order for it to be just. Those that I have called 'thoroughgoing' expose every element of any moral conception that might be implemented to the actual contractualist test. Others expose only aspects of any moral conception to the test, or expose only those moral conceptions which deviate from some default moral

conception which is not itself exposed to the test. I have suggested that thoroughgoing actual contractualism has a particular attraction distinct from that of less thoroughgoing versions, but I have not as yet offered any rationale for it. I have merely denied that the obvious rationale for less thoroughgoing versions—that they offer a way to preserve individual autonomy, understood as ‘autonomous living’—is adequate to justify the more thoroughgoing.

Modal contractualists suppose that the ideal of individualistic justification is satisfied not when every relevant individual actually does accept the moral conception in question but when everyone *should*. What makes this an interpretation of the ideal of individualistic justification rather than merely the view that what is right is that which is simply best supported by reason is, I have argued, the modal contractualist’s respect for the Individualist Restriction, according to which the reasons that explain why everyone should accept the moral conception in question must cash out in terms of the implications of that and competing moral conceptions’ implications for single individuals. This is what is distinctive about modal contractualism.

I have suggested that modal and actual contractualism might be combined in a ‘hybrid theory’. In a hybrid theory, a modal contractualist framework—the acceptance of the Individualist Restriction—is allied to the stipulation that actual non-acceptance constitutes (sufficient) grounds for reasonable rejection. This is a *prima facie* attractive view, since it combines the modal contractualist’s plausible interpretation of the contractualist ideal (see Section 2 above) with the attraction of the actual contractualist veto for each individual. Moreover it gives substance to the idea that justice proceeds from a reconciliation of the interests of reasonable individuals rather than from a mere balancing of each person’s interests, reasonable or not. But I argued that anyone hoping to advance a hybrid theory must offer a

solution to the puzzles I outlined in Section 6. They must show why it does not have arbitrary content, and they must vindicate the assumption that it incorporates the attractions not only of modal contractualism but also of actual contractualism.

The contractualist frameworks thus laid out and the hybrid theorist's challenge posed, I turn now to the development of a hybrid theory based on Rawls's views in *PL*. It is actual contractualist in a way that is consistent with modal contractualism. My reading of Rawls combines, I hope, the attractions of actual and modal contractualism, and vindicates the view that Rawls is more of an actual contractualist than has been standardly supposed.

**PART II: RAWLS AND
HYBRID
CONTRACTUALISM**

Chapter 3: Rawlsian Contractualism

1. Introduction	112
2. Political Liberalism and hybrid contractualism	114
3. The burdens of judgment as a basis for actual contractualism	127
4. The hybrid view and the actual contractualist motivation	140
5. The hybrid view and acceptance	152

1. *Introduction*

Thus far in this thesis I have considered two kinds of contractualist approach, actual contractualist and modal contractualist, and I have suggested the possibility of a third, ‘hybrid contractualism’, which combines the attractions of those two. I now want to turn to the task of elaborating and developing a hybrid view based on John Rawls’s views as set out in *Political Liberalism* (hereafter *PL*) and *Justice as Fairness: A Restatement* (*J4F*). Certain features of Rawls’s original position, as expounded in *A Theory of Justice* (*Theory*), led Scanlon in “Contractualism and Utilitarianism” to

propose/understanding Rawls as what I have been calling a modal contractualist.

Since then many have supposed that this is an accurate characterisation, and this has informed readings of the view Rawls advances in *PL*. In this chapter, however, I want to argue that Rawlsian contractualism should not be read as straightforwardly modal contractualist in the way that Scanlon proposes. In the first place, on Rawls’s own view it includes a significant, though circumscribed, commitment to actual contractualism as well as an account of the fundamental basis of the actual contractualist interpretation of the individualistic contractualist ideal. This makes it

plausible to suppose that Rawls is proposing a qualified form of hybrid contractualism. (The qualification makes the hybrid view also a mixed condition view [see Chapter 1, section 2 above]: the actual contractualist element creates a test only for some parts of a just society's moral conception.) This should be viewed as a good thing, since as I have claimed already, hybrid theories combine the attractions of both interpretations of the contractualist ideal. We can draw on Rawls's account, moreover, to explain the attraction of thoroughgoing actual contractualism and make good the lack that I highlighted in Chapter 1. In the second place, I offer an objection to the reasoning that leads to the circumscription of the actual contractualist element in Rawls. I then reconstruct his view in the light of this objection. The resultant hybrid theory is thoroughgoing in its actual contractualism and founded on a modal contractualist framework. It is a dual condition view. For the remainder of this thesis I elaborate this approach and try to show that Rawls's arguments for the qualified hybrid contractualism from which it is developed can also be employed in its defence.

I begin here with a reading of *PL* which I take to capture what Rawls intended, / although I bring out elements of the view thus revealed which are not emphasised by Rawls himself. I argue that this reading shows that Rawls should be thought of as a qualified hybrid contractualist in the way I described above. I then challenge the basis of the qualification, namely the claim that acknowledgment of the 'burdens of judgment', which explain the reasonableness of disagreement on fundamental moral matters, entails refusal to endorse moral conceptions that don't safeguard the basic liberties. That clears the way for a reconstruction of Rawls's hybrid contractualism as including a thoroughgoing actual contractualist test.

I turn next to a discussion of the puzzles that face hybrid contractualism which I raised at the end of Chapter 2. I suggest a solution to the indeterminacy puzzle

which also permits us to sidestep the scope puzzle. The solution is that we should not see being reasonable, from the point of view of the hybrid theory's modal contractualist framework, as involving the absolute prioritisation of the reason that modal contractualists claim we have to find and comply with principles that each person actually accepts. Instead we should see it as involving a readiness to reconsider one's own view in the light of differences with others. This permits us to sidestep the scope puzzle because it entails that even unreasonable individuals' acceptance matters. For we cannot suppose that if these unreasonable individuals were reasonable they would not continue to affirm the same moral conclusions anyway. So what it is reasonable to expect them to accept is not necessarily different from what they actually do accept. Since the hybrid theorist's modal contractualist framework involves asking what it would be reasonable to expect even unreasonable individuals to accept, that framework doesn't restrict the actual contractualist element's scope any further.

For the remainder of the chapter I consider the question whether what the burdens-of-judgment basis for the Rawlsian hybrid theory requires—non-conflict between an individual's moral outlook and the moral conception we seek to impose—can be the foundation for an adequate construal of actual acceptance. I claim that it can.

2. *Political Liberalism and hybrid contractualism*

PL proposes a 'political conception' of justice as fairness, the (in my terms) moral conception that Rawls defended in *Theory*. What makes it a political conception, Rawls explains, is that it is limited in applicability to the basic structure, presented

with no reference to the non-political values of any comprehensive doctrine (it has what Raz calls ‘shallow foundations’)¹⁰¹, and constructed using the strategy of ‘political constructivism’ from fundamental ideas which are shared by the members of a democratic society as part of their public political culture.^{102,103} These features make it ‘freestanding’. Justice as fairness as presented in *Theory*, on the other hand, is said to be extendable in principle to cover more than the basic structure,¹⁰⁴

¹⁰¹ See Raz, “Facing Diversity: The Case of Epistemic Abstinence”.

¹⁰² See *PL*, pp. 11-5. Rawls explains the idea of a comprehensive conception or doctrine as follows: “A moral conception is general if it applies to a wide range of subjects, and in the limit to all subjects universally. It is comprehensive when it 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...our life as a whole. A conception is fully comprehensive if it covers all recognized values and virtues within one rather precisely articulated system; whereas a conception is only partially comprehensive when it comprises a number of, but by no means all, nonpolitical values and virtues and is rather loosely articulated. Many religious and philosophical doctrines aspire to be both general and comprehensive” (*PL*, p. 13).

¹⁰³ Rawls also says that the political presentation of justice as fairness takes no stand on the truth of the principles in which it issues (in Raz’s terms, he is committed to ‘epistemic abstinence’). He seems to think that this is important because each comprehensive doctrine has an “associated theory of truth and the status of values” (*PL*, p. 150). The implication is that asserting the truth of the political conception must involve drawing on a “theory of truth and the status of values” which will be associated with a comprehensive doctrine. This idea, which we can call the doctrine of the embeddedness of truth, seems to me to be at best dubious, and in any case, as Raz argues, it is implausible to suppose that we can advocate employing principles without committing to their truth. So I ignore epistemic abstinence here.

¹⁰⁴ See *Theory*, p. 111/95.

justified partly by reference to a Kantian comprehensive doctrine,¹⁰⁵ and as such constructed using ideas which would probably not form part of a liberal democratic society's public political culture, which would be pluralistic (i.e. characterised by its members' adherence to a range of comprehensive doctrines).¹⁰⁶ Rawls hopes to employ justice as fairness as a political conception in order to avoid a problem that he identifies in the argument of *Theory*: that the Kantianism involved in justice as fairness's account of stability makes it "inconsistent with realizing its own principles under the best of foreseeable conditions." It is inconsistent in this way because the liberal institutions of justice as fairness would themselves *encourage* a pluralistic society, and in such a society it wouldn't be reasonable to expect each person to accept the Kantian justification of those institutions.¹⁰⁷

Among the fundamental ideas out of which the political conception is constructed is the idea of society as a fair system of cooperation,¹⁰⁸ which brings in its train the idea of citizens as capable of cooperating on fair terms,¹⁰⁹ which in turn involves the idea of citizens as *reasonable* in a particular sense which Rawls specifies in detail.¹¹⁰ These ideas together form 'the object conception' of society as a fair system of cooperation between citizens conceived as free, equal, reasonable, and

¹⁰⁵ The appeal to the Kantian doctrine, as Rawls sees it, is a component in the argument for 'congruence', according to which individuals will see the sense of justice as part of their own good. That argument concludes at *Theory*, pp. 574-5/503, but stretches across the final two thirds of the book. See pp. 252-3, 222, 445, 390, 515, 452, 528/462-3, 561, 491, 563, 493, 572/501, and 574-5/503. For helpful exegesis see Samuel Freeman, "Congruence and the Good of Justice". ✓

¹⁰⁶ See *PL*, p. xviii.

¹⁰⁷ *PL*, p. xix. I say more about this below.

¹⁰⁸ *PL*, p. 15.

¹⁰⁹ *PL*, pp. 16-20, especially pp. 18-9.

¹¹⁰ See *PL*, pp. 19, 48-54.

rational. Reasonableness is characterised by two basic aspects. The first aspect of the reasonable involves a desire to cooperate on and readiness to propose fair terms of cooperation that it is reasonable to expect all to accept. (This is modelled in the political conception of justice as fairness by the employment of the original position, with all the conditions it imposes upon the parties' reasoning, and the priority of the principles it produces as those appropriate to regulate social cooperation.)¹¹¹ The second aspect of the fundamental idea of the reasonable involves acknowledgment of what Rawls calls 'the burdens of judgment'.¹¹² The burdens of judgment tell us that, and why, it's reasonable for people who are themselves reasonable and rational to differ even in ideal conditions on the fundamental moral issues which are the stuff of comprehensive doctrines, so that none is at fault. (This, too, could be said to be modelled by the use of the original position in the following sense: the original position abstracts from individuals' comprehensive conceptions of the good, and so captures the idea that it is unreasonable to set things up so that any comprehensive conception operates as the standpoint from which we work out principles of justice.)¹¹³ In fact, Rawls's discussion makes it clear that the burdens of judgment explain more than that. They explain why it's reasonable to differ on moral matters *simpliciter* (i.e. not just those involved in comprehensive doctrines) as well as on some empirical and theoretical matters. I'll say more about this in the next section.

According to Rawls, acknowledging the burdens of judgment sets limits on what can be reasonably be justified to others. It is therefore in general unreasonable to impose a moral conception on another who does not accept the principles appealed

¹¹¹ *PL*, pp. 49-50, 77-8, 305-7.

¹¹² *PL*, II, §3.

¹¹³ See *PL*, pp. 61-2, 305.

to in its justification, since it is not unreasonable of her to reject them. However, one thing that the burdens of judgment show us *can* reasonably be justified to and so imposed on others are the basic liberties. I'll discuss Rawls's reasons for supposing this in more detail in the next section too.¹¹⁴

Now the question of the stability of justice as fairness as a political conception is this: can this freestanding conception "gain sufficient support" in the society it regulates?¹¹⁵ But as it transpires, Rawls is not interested in just *anyone's* support. He is interested in the support that is provided by an overlapping consensus of *reasonable* comprehensive doctrines.¹¹⁶ Whether or not there exist enough adherents of unreasonable doctrines which do not support the political conception to undermine stability *simpliciter* is a matter which we leave aside. The real aim is to show that each person could reasonably be expected to endorse the principles of justice.¹¹⁷ Only then will stability be "stability for the right reasons".¹¹⁸

The condition that each person 'could reasonably be expected to endorse' the political conception is a modal contractualist condition. In seeking an overlapping consensus of reasonable doctrines Rawls takes himself to be seeking a consensus of individuals who are reasonable. The two aspects of the reasonable he explicitly connects with Scanlon's contractualist account of motivation. Scanlon says that the basic contractualist motivation is a desire to respond to the reason that we have to

¹¹⁴ *PL*, p. 61.

¹¹⁵ *PL*, p. 65.

¹¹⁶ See *PL*, p. 65 and e.g. pp. 134, 143-4.

¹¹⁷ *PL*, p. 13⁷. Cf. p. xliv.

¹¹⁸ *PL*, p. xxxix, n. 5. Cf. p. 391, n. 2⁷.

✓ but Rawls' ~~just~~ ^{desire} to do this
expected from within a contractual framework.

justify our actions to others on grounds that they could not reasonably reject.¹¹⁹

Rawls says that desiring to cooperate on and being ready to propose fair terms of cooperation that it is reasonable to expect all to accept and accepting the consequences of the burdens of judgment are “two related expressions of [the Scanlonian] desire”.¹²⁰ So he sees himself as seeking an overlapping consensus among those who respond appropriately (i.e. by having the Scanlonian desire) to the Scanlonian reason that they have to find and act on principles that no one could reasonably reject. The ‘reasonable to expect others to accept’ formula should be seen as synonymous with the reasonable rejection formula. (Scanlon himself suggests that the latter is merely a more precise version of the former.)¹²¹

Rawls generally speaks of an overlapping consensus of reasonable *doctrines*. So we should make clear the connection between doctrines and individuals.¹²²

Reasonable comprehensive doctrines are, roughly speaking, more or less coherent normative systems expressing a comprehensive and intelligible view of the world

¹¹⁹ See *WWO*, pp. 153-8. In “Contractualism and Utilitarianism”, which is the text that Rawls cites, Scanlon refers to a desire to justify our actions to others on grounds that they could not reasonably reject. In *WWO*, Scanlon argues that what’s basic is the reason that we have to do so, not the desire to do so. See pp. 7-8.

¹²⁰ *PL*, p. 49, n. 2.

¹²¹ See *What We Owe To Each Other*, p. 4.

¹²² At “John Rawls and the Search for Stability”, p. 898, Barry notes that it is “people, not doctrines, that go around endorsing conceptions”. (Cf. Scheffler, “The Appeal of Political Liberalism”, p. 11.) I think Rawls’s talk of doctrines affirming the political conception arises innocently out of a wish to stress that individuals’ endorsement of the political conception is based on the grounds provided by their comprehensive doctrines. Describing things this way avoids giving the impression that individuals’ support for the political conception is instrumental in the way that it is in a *modus vivendi*. See below.

and which normally belong to a tradition of thought and doctrine.¹²³ More importantly, though, what makes a comprehensive doctrine unreasonable is conflict with aspects of the reasonable itself: “[w]e avoid excluding doctrines as

unreasonable without strong grounds based on clear aspects of the reasonable itself.”¹²⁴ So, for example, one way in which a doctrine would be unreasonable is if it refused to acknowledge the burdens of judgment. Reasonable comprehensive doctrines, then, are those familiar comprehensive doctrines whose adherents also recognise the burdens of judgment and their consequence, the basic liberties; and who have the desire to cooperate on fair, reasonable terms which is the first aspect of reasonableness. The reasonableness of individuals is basic here, since it is primarily by reference to the aspects of reasonableness of individuals that we determine the reasonableness of a doctrine.

So what's needed in order to show that it's reasonable to expect each reasonable person to endorse the political conception of justice as fairness? Well, we know (from the second aspect of the reasonable) that it's unreasonable to expect *anyone* to affirm any particular comprehensive doctrine (at least that they don't already affirm), though it's reasonable to expect them to affirm at minimum the basic liberties as part of their conception of justice. We know (from the first aspect) that it's unreasonable not to be willing to propose and abide by fair terms of cooperation that it's reasonable to expect others to endorse. And we know that we're faced with a pluralistic society. So we can say at least this: what we're looking for is a liberal conception (this is what I'll call a conception that safeguards the basic liberties)

? / that's not based only on a single comprehensive doctrine. / That means that we need

¹²³ *PL*, p. 59.

¹²⁴ *PL*, pp. 59-60.

a freestanding liberal political conception. Moreover, we know that everyone shares¹²⁵ the various conceptions of persons and society that the political conception draws together.¹²⁶ So it would be reasonable to expect everyone's endorsement of a freestanding liberal conception developed from *these*—which is precisely what justice as fairness, in its political presentation, is. (If the fact that everyone shares them doesn't make it reasonable to expect them to affirm them, then it seems that we can't go any further than the basic liberties.) And that, surely, is in fact all we need. Of course, from the fact that this is what it would be reasonable to expect *everyone* to endorse we can derive the conclusion that there will be an overlapping consensus among *reasonable* people—those who, like us, are interested in what it's reasonable to expect everyone to accept. But this seems to be an unnecessary further move. Isn't the point that it's reasonable to expect everyone to accept justice as fairness, whether or not there is an overlapping consensus on it?

Yet Rawls expends a good deal of effort on discussion of the overlapping consensus. Why? One explanation is that showing that it would be unreasonable to reject justice as fairness, given what Rawls takes as given (the definition of the reasonable, the shared elements implicit in the public culture), is not showing that enough people would be reasonable to render justice as fairness stable. If it is stable, it's stable for the right reasons—but how do we know it's stable? However, the stability question can be answered using the psychology of moral development that Rawls offers in *Theory* and merely rubber-stamps in *PL*.¹²⁷ We don't need anything more for that. A better explanation is that the characterisation of a

¹²⁵ *PL*, pp. 13-4.

¹²⁶ *PL*, p. 304; see also “Kantian Constructivism in Moral Theory”.

¹²⁷ *PL*, p. 143, n. 9.

reasonable, political conception-endorsing person needs moral-psychological plausibility. This we get from Rawls's account of how reasonable individuals' wider comprehensive doctrines connect up with the narrower realm of the values of the (freestanding) political conception. First, we are told, the values articulated by the political conception are very important.¹²⁸ And second, they're congruent with (that is, supported by or at least not inconsistent with) a variety of non-political values¹²⁹ —the overlapping consensus is not a *modus vivendi*.¹³⁰ The first part of this account seems to be of a piece with the basis for the liberal principle of legitimacy, viz., that political power is coercive and (more or less) inescapable.¹³¹ The thought is that the political sphere has a deep importance which translates into priority for the values appropriate to it. The second part relies for its plausibility on Rawls's rather brief account in Lecture IV, §8 of the relations between various comprehensive doctrines in his 'model case'¹³² and the political conception, which purports to demonstrate that affirming the values of the political conception is "not a compromise" for those who also affirm the comprehensive doctrines of the model case.¹³³

¹²⁸ *PL*, p. 139.

¹²⁹ *PL*, p. 140.

¹³⁰ See *PL*, pp. 145-9.

¹³¹ See *PL*, p. 137.

¹³² *PL*, pp. 145-6.

¹³³ *PL*, p. 170. See also pp. 385ff. The brevity of this account might be explained to some extent by Rawls claim that we must simply leave individuals to work out for themselves how the political conception fits with their wider views (see *PL*, p. 140). *Some* kind of account is necessary, however, in order to lend plausibility to the idea that it's even possible for them to do this. See Chapter 4 below.

But the most important explanation for Rawls's focus on the overlapping consensus, I think, is that he wants not merely to assert that but also to illustrate the way in which support for political liberalism leaves room for a variety of accounts of the realms of non-political values, in contrast with other approaches to justice which are based on a single comprehensive doctrine (such as the Kantian Interpretation of justice as fairness which Rawls presents in *Theory*). Describing how different people reconcile the political conception with their comprehensive doctrines shows us how what's demanded of individuals by reasonableness is compatible with the range of wider moral views which we'd find in a pluralistic liberal society. And why, we might ask, is it important to leave room for a variety of accounts of the realms of non-political values? One reason is purely practical: given the desideratum of full publicity (in which a full argument for a conception of justice is made publicly available), stability is harder to secure in a liberal, pluralistic society if there *isn't* room left by that argument for the variety of accounts. But it's only harder, which doesn't entail that it's very hard; and anyway if we suppose that we've got the conception of justice right in the first place, it would be odd to insist on full publicity at the cost of any possibility of achieving justice. The more important reason is the reasonableness of differing on fundamental moral matters (the consequence of the burdens of judgment) which makes it unreasonable to expect everyone to accept any justification based on fundamental moral commitments that they don't actually share, except if these fall out from acceptance of the burdens of judgment themselves. So it's important to show—and not merely assert—that the political conception doesn't involve any such justifications—that it doesn't contradict any reasonable individual's views. Hence we try to show that justice as fairness is capable of support from a range of moral positions: an overlapping consensus.

Strictly speaking, the *actual existence* of such a consensus is not a condition of stability for the right reasons. All we need is stability and a political conception which it's reasonable to expect each person to accept—even if no one does. It might be that no one is reasonable, yet sufficient numbers for stability endorse justice as fairness anyway. But if two or more people *are* reasonable, an overlapping consensus is necessarily part of stability for the right reasons for the following reason.¹³⁴ It is a crucial part in the argument for the reasonableness of expecting acceptance of justice as fairness that the political conception is built out of *shared* elements. Justice as fairness, and not some other liberal political conception (recall that all reasonable individuals affirm the basic liberties as a consequence of the burdens of judgment), is appropriate because it elaborates these elements which everyone shares. It would be reasonable to expect acceptance of some other liberal political conception if we took other political ideas to be shared.¹³⁵ Since everyone shares them, including anyone who is reasonable, it is reasonable to expect everyone to endorse justice as fairness. And since reasonable people are reasonable, this is what they will do. So there will be an overlapping consensus of reasonable people if there *are* two or more reasonable people.

What this brings out is the importance of the role that people's *actual acceptance* of these shared elements plays. I can't reasonably expect you or anyone else to endorse any conception of justice (the basis of) which conflicts with your comprehensive doctrine. And since there are lots of different and conflicting comprehensive doctrines around, we need to come up with a freestanding political

¹³⁴ I say 'two or more' because of course one person's acceptance doesn't constitute a consensus. But obviously if there's one person, her endorsement of the political conception is a necessary condition of stability for the right reasons for the reasons I go on to give.

¹³⁵ *PL*, p. 167.

112 *replies* *that* *can* *be* *related* *to* *the* *Rawls*
doctrine, *and* *does* *not* *apply* *to*
any *doctrine* *or* *any* *political* *doctrine*

conception which doesn't draw on any single comprehensive doctrine, at least insofar as it conflicts with others. But notice that the political conception needs to be congruent with the *political* parts of each person's moral outlook, too. For the burdens of judgment, which determine this requirement in combination with the first aspect of the reasonable, don't distinguish between political values and wider values: it's unreasonable to expect anyone (reasonable) who differs from you in their view to accept yours on important moral matters, political and otherwise, with the exception of endorsement of a liberal conception of justice. So if we *can* come up with a political conception which doesn't conflict with the moral outlook of anyone reasonable, that's because everyone *just does agree* on the fundaments of such a conception, not because the political realm is exempt from the burdens of judgment (liberal conceptions aside) so that it's reasonable to require agreement on those fundaments. In other words, it's because it passes an actual contractualist test. This is Rawls's actual contractualism.

It is not the case, of course, that each person in the overlapping consensus actually accepts the relevant political conception in the sense that she actually (explicitly or tacitly) consents to it. Nor is it even necessarily true that she *would* consent to it if she considered the question (that depends on the conditions invoked in the hypothetical). *What* matters is that a person's actual moral commitments support, or at least do not conflict with, the political conception. But as I shall argue at the end of this chapter, that is sufficient for this part of Rawls's contractualism to be counted *actual* contractualism. Congruence of moral commitments with the political conception is a viable actual contractualist construal of the variable of acceptance. //

So the burdens of judgment are absolutely central to political liberalism. The 'double role' that Rawls ascribes to them in his introduction to the paperback

edition of *PL* seriously underplays their significance, though it is correct that they have a double role. There, Rawls claims that “they are part of the basis for liberty of conscience and freedom of thought founded on the idea of the reasonable. And they lead us to recognize that there are different and incompatible liberal conceptions”.¹³⁶ The first half of this is right, although it doesn’t bring home the force of the idea that it’s unreasonable by Rawls’s lights not to accept a liberal conception. But the second half strikes me as odd. Who would have supposed that there *weren’t* different and incompatible liberal conceptions? Moreover, this doesn’t bring out at all the actual contractualist force of the burdens of judgment, even in the restricted sense that applies if we accept the first part. The burdens of judgment make it unreasonable to expect anyone to endorse anything more specific than this unless they actually already *do* endorse it in some sense.

This makes political liberalism at once easy and difficult to achieve. On the one hand, it seems to get us to liberalism (i.e., the basic liberties) with barely any effort. Brian Barry complains, indeed, that the burdens of judgment argument for liberalism obviates the need for an overlapping consensus.¹³⁷ On the other hand, to get from liberalism to specifically justice as fairness (i.e., from the basic liberties to these plus Rawls’s second principle of justice) requires that justice as fairness pass an actual contractualist test: we need everyone reasonable¹³⁸ to endorse the

¹³⁶ *PL*, p. xl ix.

¹³⁷ See Barry, “John Rawls and the Search for Stability”, pp. 898-901.

¹³⁸ We also need to know what could be reasonably expected of unreasonable people, people who fail to recognise the modal contractualist reason that they have to find and act on principles that no one could reasonably reject or who fail to acknowledge the burdens of judgment. For the modal contractualist framework requires that *no one* could reasonably reject a just moral conception, not merely that no one who is as a matter of fact reasonable could reasonably reject a just moral

fundamental ideas from which the political presentation of justice as fairness is elaborated. *Prima facie* this is not an easy test to pass.

We have seen that Rawls operates, in *PL*, within a modal contractualist framework. We have also seen that he endorses a (somewhat circumscribed) actual contractualist test. As I said in the introduction to this chapter, this is part of a mixed condition actual contractualism. It only applies to elements within a liberal framework, which itself is not subject to any actual contractualist test but entirely accounted for by the modal contractualist framework, which identifies the burdens of judgment as a consideration in favour of a liberal conception. In the next section I shall challenge this part of Rawls's view. The basis of the actual contractualist test, meanwhile, is also given by the modal contractualist framework: given the burdens of judgment, it's *reasonable to reject* a specific liberal political conception on grounds of actual non-acceptance. This basis for the actual contractualist test shows that Rawls should be viewed as a hybrid contractualist.

3. *The burdens of judgment as a basis for actual contractualism*

As I noted above, the burdens of judgment explain how people can disagree about matters of morality and the good without either party to the disagreement being at fault. I want now to discuss Rawls's reasoning about the burdens of judgment in

conception. So we need to know what it would be reasonable for unreasonable individuals to reject. Since it is reasonable, if you're reasonable, to reject a liberal conception that you don't actually accept, we need to know what these people would accept if they were reasonable so that we can find out what they wouldn't. This has implications for the puzzle of scope reconciliation for hybrid theorists that I raised at the end of Chapter 2. I discuss it later on in this chapter.

more detail. Obviously the cause of some disagreements is that one side *is* at fault. People reason badly, or they fail to see relevant considerations, or they simply make mistakes. And in other moral and ethical disagreements *both* sides are at fault for the same reasons. But Rawls supposes that these are not the only explanations of disagreement about such matters. Sometimes individuals disagree without anyone's being unreasonable or ~~having made a mistake~~. This is because of "the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment"¹³⁹—the many ways in which two reasonable people using correctly and conscientiously their equal powers of reason and judgment can arrive at different conclusions in moral (and other) matters.

What are these hazards? Most of those that Rawls lists (although he says that there are others) refer to the difficulty that we have in comprehending and appropriately weighting the full range of considerations which bear upon our moral and ethical judgments. The evidence is "conflicting and complex"; many ways of weighing it are reasonable; our concepts are "vague and subject to hard cases"; we cannot tell how far our judgments are "shaped by...our whole course of life up to now"; and it is difficult to make an assessment of the "different kinds of normative consideration of different force on both sides of an issue".¹⁴⁰ Moreover, not every value, even if it is agreed by all to *be* a value, can be realized by one society.¹⁴¹ What is essential to understanding how these factors explain *no-fault* disagreement is that there is no method by which we may be sure that we have come to the right assessment. Our concepts, for example, just are vague; it's not that some people are

¹³⁹ *PL*, p. 56.

¹⁴⁰ *PL*, p. 57.

¹⁴¹ As Rawls notes, this is a theme of Isaiah Berlin's; since *PL* it has also been one of Rawls's.

less vague in their concepts than others. The evidence itself is conflicting, not just our interpretations of it. The burdens of judgment are significant because together they show us that the subject matter of morality itself involves uncertain cases and unclear boundaries, that the moral landscape is in itself blurry. The problem is not just with our means of accessing it. As a result, disagreement persists, and is reasonable, even when the best means of accessing the issues are used.

To acknowledge the burdens of judgment, then, is to acknowledge that reasonable people may well differ epistemically blamelessly on even the most important of moral and ethical judgments. (Note that Rawls's description of the burdens explains not only disagreement on comprehensive moral issues but also on almost all moral and ethical—including political—issues. The qualification arises from the putative indisputability of the burdens of judgment themselves and the existence of the modal contractualist reason.) Thus the fact of pluralism is almost inevitable in a free society: individuals will come to affirm different comprehensive doctrines.¹⁴² Reasonable people see that reasonable others could blamelessly fail to reach the same conclusions as them and recognise therefore that others will not necessarily accept their comprehensive doctrine-based justifications of the use of state power.

That they will not necessarily accept such justifications is one thing. What to do about that fact is another. This is where Rawls makes his distinctive move: he identifies a reasonable individual's non-acceptance of some comprehensive doctrine-based justification—on the grounds of its conflict with the view that she, not unreasonably, takes of the moral landscape—as grounds for denying that she could reasonably be expected to accept it. The mere fact of her non-acceptance is

¹⁴² See *PL*, p. 60.

blameless

enough once we "acknowledge" the burdens of judgment. In other words, Rawls identifies the mere fact of non-acceptance as (effectively) a reason in the modal contractualist framework for the rejection of a moral conception. He writes:

Since many doctrines are seen to be reasonable, those who insist, when fundamental political questions are at stake, on what they take as true but others do not, seem to others simply to insist on their own beliefs when they have the political power to do so. Of course, those who do insist on their own beliefs also insist that their beliefs alone are true: they impose their beliefs because, they say, their beliefs are true and not because they are their beliefs. But this is a claim that all equally could make; it is also a claim that cannot be made good by anyone to citizens generally. So, when we make such claims others, who are themselves reasonable, must count us unreasonable.¹⁴³

The key thought in this passage is that it would be unreasonable to impose what I am calling a moral conception (i.e. in the political realm) on the basis of the truth of a comprehensive doctrine which supports it. Importantly, this is not supposed to be because that doctrine is *not* true, or even because we cannot *know* that it is true—Rawls denies that he is offering a sceptical view of the sorts of claims in question¹⁴⁴—but rather because others might reasonably affirm different comprehensive doctrines and therefore reasonably disagree about the truth of the doctrine in question. It would be unreasonable to impose the moral conception on

¹⁴³ *PL*, p. 61. Rawls goes on to conclude it is unreasonable not to endorse some form of liberty of conscience and freedom of thought. I don't see that this follows: see below.

¹⁴⁴ See *PL*, pp. 62-3, 138, and 150-154.

the basis of the truth of one's comprehensive doctrine because it would be reasonable for others not to accept that comprehensive doctrine. The reasonableness of disagreement is the fundamental fact.

The burdens of judgment can help to explain the attraction of what I called 'thoroughgoing' actual contractualism—at least within the modal contractualist framework. In Chapter 1 I argued that the attraction of actual contractualist views which submit almost every aspect of proposed moral conceptions to the actual contractualist test cannot adequately be explained in terms of a fundamental commitment to the ideal of autonomous living. It can, however, be explained in terms of a conception of individuals as equally authoritative in their epistemic access to the moral landscape yet reasonably differing in their judgments about it. We should care about actual acceptance, even of principles and norms which determine the degree to which we lead autonomous lives in the sense which is often invoked in support of liberal moral conceptions, because without it we have no way to justify the imposition of those norms to others given that we ourselves are no more authoritative in our judgment about the moral landscape than those upon whom we seek to impose those norms. Our *own* acceptance of them is sufficient for the legitimacy of their imposition upon *us*, since our acceptance of them constitutes acceptance of their truth and their truth would constitute a justification for imposing them. But the reasonableness of not accepting them entails the unreasonableness of expecting someone who does not accept them to accept a justification for their imposition which is based upon their truth. In the context of a conception of individuals as equally authoritative and reasonable in differing in their judgments about the moral landscape, an appeal to the truth of a moral conception, even if it *is* true, does not work as a justification that others may reasonably be expected to accept—unless everyone already does actually accept it (that it is true).

In that case it is reasonable to expect each person to accept, by way of justification, an appeal to the truth of the moral conception. But since it would be reasonable for anyone to deny the truth of the moral conception, such an appeal is effectively an appeal to the fact that she *accepts* its truth.

Consider by way of analogy the following example. You and I are purchasers for an art museum. We have a limited budget, enough for one of two paintings that are on offer. Neither is historically more significant than the other or more suited to our present collection: our decision will be on aesthetic grounds. The painting that we don't choose will be bought by a private collection and we will not have a chance to buy it again. I think that we should purchase the Flemish landscape, which I think (correctly, let us say) is aesthetically superior to the Italian portrait. In making a case for it to you, I could appeal directly to the truth of its aesthetic superiority. I could insist that we purchase it on this basis. But it is reasonable to disagree about this, at least when these two paintings are in question // so to insist on purchasing it on the basis of its aesthetic superiority *simpliciter* would be unreasonable. // I recognise this, so although I may still appeal to the truth of its aesthetic superiority, I do so conscious that my case for it has no force at all if you don't accept that it *is* aesthetically superior. I can, of course, point out my reasons for taking it to be so. But when all is said and done an appeal to the Flemish landscape's aesthetic superiority as grounds for its purchase must be, to justify its purchase to you, an appeal to *your acceptance* of its superiority. I don't say: "it's aesthetically superior—so let's buy it." I say: "*we're agreed that* it's aesthetically superior—so let's buy it."

The burdens of judgment make it appropriate to treat moral judgments like the aesthetic judgments in this example. This is not to say that moral judgments are not true or false. It is to say that it is unreasonable to appeal to their truth and

How did we get the beautiful colored trees
& such vegetation as this?

falsehood to justify the imposition of a moral conception. Instead, we must appeal to the fact of individuals' actual endorsement of them. Each person's actual endorsement of a moral conception (or a basis for that moral conception—see the final section of this chapter) is essential to justify its imposition, and it is to that endorsement that we must appeal in justifying its imposition to them. To do otherwise is unreasonable in the light of the burdens of judgment. So here we have a burdens-of-judgment basis for actual contractualism.

I said that from her *own* point of view, a person may see the imposition of a moral conception as justified because of its truth. What the burdens of judgment suggest is that others' justifications *to* her of the imposition (i.e. society's imposition) of that moral conception upon her can nevertheless reasonably appeal only to the fact of her actual acceptance. It is sometimes suggested that left-wingers are perfectly welcome, in a low tax society, to contract themselves into paying higher taxes, if they think that such taxes ought to be imposed. So I might sign such a contract. I could then see the higher taxes imposed upon me as justified because they accord with principles I take to be true, yet at the same time refuse to accept your appeal to their truth, rather than to the fact that I signed the contract, as your justification for forcing me to pay them. I can consistently refuse to accept as a justification from you what I accept as a justification from myself. The same is true in respect of the imposition of moral conceptions. The burdens of judgment, in the modal contractualist framework, help to show us why.¹⁴⁵

¹⁴⁵ Cf. Stephen Darwall, "Contractualism, Root and Branch: A Review Essay", especially pp. 204-5, where Darwall emphasises the contrast between justifying something *to* someone and justifying something in someone's presence.

X they don't we just a step further. What the
type of just & I mean a more or less

As we saw above, Rawls does not invoke the burdens of judgment to support a thoroughgoing actual contractualism. Instead, by way of conclusion to the passage

that I quoted on p. 130 above, he claims that "reasonable persons see that the burdens of judgment set limits on what can be reasonably justified to others, and so

they endorse some form of liberty of conscience and freedom of thought."¹⁴⁶ This

restricts the range of the actual contractualist test in Rawls's view. But here Rawls

is, I think, mistaken. Liberty of conscience and freedom of thought are

components in moral conceptions that are no less subject than any other to the

difficulties raised by the burdens of judgment. The claim that these are just, like the

claim that *extra ecclesiam nulla salus*,¹⁴⁷ is "a claim that all equally could make; it is also

a claim that cannot be made good by anyone to citizens generally".¹⁴⁸ What would

be required to give liberty of conscience and freedom privileged status would be

some internal connection with or entailment by acknowledgment of the burdens of

judgment. But acknowledging the burdens of judgment is merely acknowledging

that reasonable people can disagree with in their moral views without its being the

case that there must be a mistake in the reasoning by which they came to hold

opposing positions. (And it's not the case that the basic liberties must be affirmed

in virtue simply of the *first* aspect of the reasonable.) You could consistently

acknowledge the actual contractualist test that I claimed a moment ago is grounded

in acknowledgment of the burdens of judgment and yet deny that others should be

free to act upon their opposing views. Instead, you would suppose that the

question which view can reasonably determine how people may justly act can be

146

¹⁴⁶ *PL*, p. 61.

¹⁴⁷ "Outwith the church there is no salvation". See *PL*, p. 138.

¹⁴⁸ *PL*, p. 61 (quoted more fully above).

found out only by ascertaining which view is actually accepted by everyone. You would deny that in the absence of actual agreement that justice should give its seal of approval to *any* moral conception. The burdens of judgment explain how people can disagree without fault; they do not set out ideal epistemic conditions for moral reasoning on the basis of which we can reach the conclusion that liberal principles are just.

Nor will it do to say that since each person may reasonably differ about what her own behaviour should be, it would be unreasonable for anyone to propose moral conceptions which do not permit each individual the freedom to determine her own behaviour (consistent with the same freedom for all). The point of the idea of the burdens of judgment is that people may reasonably differ in the moral conclusions they draw; their conclusions may not have this kind of individualistic liberalism built into them. A salient fact about at least some comprehensive doctrines that have existed is that realisation or furthering of the good as they specify it involves the behaviour of not only those who adhere to the comprehensive doctrines in question but also those who don't. So, for example, some religious comprehensive doctrines view the good as requiring the behaviour of all individuals—as opposed to only those who adhere to the religion in question—to conform to the prescriptions of some central religious text. Moreover, some such doctrines do not require that the relevant behaviour must be voluntarily undertaken for the good to be secured or furthered: free faith is not necessary for the realisation of the good.¹⁴⁹ Religious regimes may sometimes make conformity

¹⁴⁹ Rawls notes at *PL*, p. 170 that he “perhaps too optimistically” supposes that “all the main historical religions admit of [an account of free faith]”. I take ‘an account of free faith’ to mean an explanation or assertion of the value or necessity of the free adoption of faith. The optimistic

with religious requirements obligatory regardless of whether a person believes in the religion or requirements in question or not. So people may reasonably differ about what their own behaviour should be; but they may also reasonably differ about what each *other* person's behaviour should be. Nothing in the idea of the burdens of judgment gives us grounds to deny this.

If Rawls were right about the conclusions licensed by the burdens of judgment, the best articulation of his view would be the hybrid view with a less thoroughgoing actual contractualist component. The burdens of judgment would provide reason to reject any moral conception which did not include provision for at least the basic liberties of conscience and thought. But he is wrong about the conclusions. The burdens of judgment alone provide no basis for the supposition that there is *any* moral conception that it would be unreasonable to affirm apart from those which fail to acknowledge the actual contractualist consequence of the burdens of judgment themselves. They nevertheless do produce that actual contractualist consequence: that it would be unreasonable to impose a moral conception on someone who did not accept it.

What Rawls supposes we seek when we seek to find a basis for the overlapping consensus is a basis for unanimous acceptance of a *specific* liberal moral conception—in Rawls's favoured case a specific liberal *egalitarian* moral conception—among individuals who are already unanimous, by definition (as reasonable), in their acceptance of the necessity of *some* liberal moral conception.

But since I deny that he is right about the conclusions licensed by the burdens of judgment, I also deny that it makes sense to care about a basis for unanimous

supposition is of course consistent with its being the case that all the main historical religions have existed and do exist in versions which do not in fact include an account of free faith.

acceptance only among those who are already unanimous in their acceptance of some liberal moral conception. Does this mean we should be looking for a basis for consensus among not only reasonable but also unreasonable individuals?

We answer this question by determining whether the burdens of judgment basis for actual contractualism provides a rationale for a restriction in scope to reasonable individuals in some sense. If liberalism follows automatically from the burdens of judgment we do have such a rationale: if your fundamental premise is something that entails acceptance of liberalism, you have no reason to care about anyone's actual acceptance who is second-aspect unreasonable—which includes anyone with illiberal political views. A theory built on a liberalism-entailing fundamental premise need not try to accommodate the actual views of those who deny the premise or the entailment, though it may perfectly well take itself to be required to accommodate the views that these people *would* accept if they *did* accept the premise and entailment. The trouble is that while such a requirement entails a liberal moral conception, to get from there to a specific moral conception requires knowledge of a hypothetical which we surely can't have. Which specific liberal moral conception would someone who is illiberal (i.e. who does not endorse the basic liberties) accept if she were liberal? Her actual views in most cases will give us no guide at all. So it makes sense to concentrate on those who are second-aspect reasonable alone, and seek an overlapping consensus among their views rather than among both their views and those that second-aspect unreasonable individuals would have if they were second-aspect reasonable. This at least partly explains Rawls's restriction in the scope of his actual contractualist element to reasonable individuals.

But I am denying that liberalism does automatically follow from the burdens of judgment, and therefore we do not have this rationale for restriction in scope to second-aspect reasonable individuals. Do we have any other? You might suppose

that since we continue to affirm the burdens of judgment and their actual contractualist consequence, we have a rationale for restricting the scope of our actual contractualism to those who also affirm the burdens of judgment and actual contractualism. Why should we care about what individuals who deny the fundamental premise of actual contractualism accept or reject? But this would be analogous to a Hobbesian's refusing to require, as necessary for her to have the political obligation, the consent of an individual who rejects the necessity of consent to political obligation. Hobbesians don't do this for the following good reason: at the level of the imposition of moral conceptions against the Hobbesian moral background, whether or not anyone accepts that moral background itself is irrelevant. We are interested in their acceptance of first-order views as articulated in moral conceptions, not their views about how or whether these moral conceptions should meet the ideal of individualistic justification. Our own second-order view, which determines the details of the moral background, itself entails no particular first-order view.¹⁵⁰ Consider by analogy the right to vote in a constitutional democracy. Suppose that someone in such a democracy believes that ideally we would have socialism implemented by a benevolent dictator. In an election she votes for the socialist candidate, though it is constitutionally impossible that the winning candidate should establish a dictatorship and end the democracy. The fact that the voter rejects the democratic system which grants each person a vote in the first place has no bearing on whether her views about socialism should be taken into account by that system. Those views are compatible with the affirmation of

¹⁵⁰ Contrast Locke, whose second-order view does entail a moral conception—that which obtains in the Lockean state of nature. This includes certain inalienable rights; so if your view leads you to reject those rights, your consent is not necessary for the justice of the imposition of a moral conception to which you did not consent.

democracy, so there is nothing offensive in their informing a person's vote. Her vote should be counted like any other; the result of the election should reflect everyone's views about the first-order policies to be implemented, not just those of people who accept the second-order democratic view about how the first-order policies are to be chosen.

Matters are different, however, if the second-order view about how first-order policies are to be chosen entails certain first-order policies. Suppose now that it is a fact that democracy is impossible unless every voter is literate. In that case, it should be constitutionally impossible in our constitutional democracy not only that a winning candidate should establish a dictatorship but also that she should implement policies which bring about less than full literacy rates. Votes for such policies should not be counted or, better, possible (because no party may propose such policies). In this case the situation is analogous to the actual contractualism which follows from a liberalism-entailing conception of the burdens of judgment. Thus for such a view we arrive at a position according to which there is a rationale, as I explained two paragraphs ago, for restricting actual contractualism's scope to those who are second-aspect reasonable.

So the revision to Rawls's hybrid view that I advocate denies that liberalism follows from acknowledgment of the burdens of judgment and it therefore relaxes the restriction in the view's actual contractualist scope to those who are second-aspect reasonable. I shall discuss first-aspect reasonableness in the next section, in the context of the puzzles for hybrid contractualism that I raised at the end of Chapter 2. Before that: is there a rationale for any *other* restriction in the revised Rawlsian view's actual contractualist scope? Well, just as the Lockean restricts his actual contractualist scope to those who are capable of the freedom that is protected by the necessity of consent for the imposition of a moral conception other than that

of the state of nature, so the revised Rawlsian can restrict her actual contractualist scope to those who are capable of the kinds of judgment about which the burdens of judgment show us that individuals may reasonably disagree. So we need not worry about the actual acceptance of those who are not capable of forming and revising their conception of the good. We can, that is, follow Rawls in assuming that everyone (that we are interested in) has his 'second moral power' and the capacities that underpin it.¹⁵¹ These include a minimum degree of rationality, theoretical ability, and self-direction.

4. *The hybrid view and the actual contractualist motivation*

The revised Rawlsian view is a hybrid theory with a more thoroughgoing actual contractualist component than Rawls's official view. It says that it is unreasonable to expect each person to accept any moral conception that she actually does not accept. I want to argue that this view is the basis for a plausible elaboration of Rawlsian egalitarianism which incorporates the attractions of both modal and actual contractualism. To do this requires me to do the following. First, I must explain how this hybrid view solves the two puzzles that I introduced at the end of the last chapter. I try to meet this requirement in the remainder of this chapter. In Chapter 4 I shall go on to connect this hybrid framework in more detail with Rawls's strategy of political constructivism, thereby explaining how we might go about showing that (what I am calling) a moral conception could be justified on the hybrid view. How should it be argued that any particular moral conception *would* actually be accepted by each relevant person in society? In the final chapters of the thesis

¹⁵¹ See *PL*, pp. 18-20, 72.

I'll elaborate and endorse Rawls's own argument that specifically the egalitarianism of justice as fairness would be actually accepted in the way necessary for it to be individualistically justified on the hybrid view.

Given the burdens of judgment as an explanation of the fundamental basis of actual contractualism, it makes sense, as we saw above, for an actual contractualist to restrict scope to those individuals who meet the following condition: they must be capable of the kinds of judgment about which individuals may reasonably disagree, given the burdens of judgment. There is nothing objectionable in an actual contractualist's restricting scope in this way when her view is founded on her acceptance of the burdens of judgment. But it would be objectionable, because incompatible with her stated motivation, for her to restrict scope in other ways. If, for example, she also excluded those individuals whose views were very right-wing, or those whose eyes were blue, her view would no longer retain the attractions of *actual contractualism* regardless of the appeal (for us, say) of the moral conception that popped out at the end. To retain those attractions a theory must remain true to actual contractualism's fundamental motivations—in this case appreciation of the burdens of judgment. The non-acceptance of those who meet the condition mentioned above, that is, must delegitimise proposed moral conceptions.

This brings up once again the second puzzle for hybrid theorists that I raised at the end of Chapter 2. There, I said that a hybrid contractualist can consistently ignore the non-acceptance of people who are unreasonable in the sense that they fail to recognise the reason that there is (that the modal contractualist element of the hybrid view takes there to be) to find and comply with a moral conception that each person actually accepts. Otherwise she seems to allow what she takes to be unjust views to contaminate the content of justice. The difficulty is this: this further restriction in scope restricts the constituency of those whose actual acceptance is

necessary for a moral conception to be just beyond the limits which follow from appreciation of the burdens of judgment, the actual contractualist motivation. So doesn't the hybrid view fail, after all, to incorporate the attraction of actual contractualism? Insofar as we are actual contractualists, shouldn't we care about acceptance of proposed moral conceptions by even those who aren't reasonable in the modal contractualist and Rawls's first sense? I called this 'the puzzle of scope reconciliation'.

However, the hybrid view can be elaborated in a way that deals with this worry. To explain how this can be done involves explaining how it can overcome the other puzzle that I outlined at the end of Chapter 2. This was the 'arbitrariness puzzle': without the introduction of factors about which their view provides no guidance, hybrid theories have no determinate outcome, and any outcome they have is from the theory's point of view arbitrary. The problem arises precisely because of the nature of the first aspect of reasonableness, necessary from a modal contractualist point of view for an individual's non-acceptance of some moral conception to give reasonable grounds for its rejection. Reasonable individuals care about finding and complying with principles or moral conceptions that everyone can reasonably be expected to accept. According to the burdens of judgment, individuals can reasonably be expected to accept only principles or moral conceptions they actually accept. It is plausible to suppose that they actually accept only what does not conflict with their own moral commitments.¹⁵² But if they are reasonable (in the first aspect) themselves—and notwithstanding the puzzle of scope reconciliation, those whose non-acceptance we have reason to care about will be—then this may make them willing to alter their moral commitments to make it possible to find

¹⁵² I say more about this in the following section.

principles that do not conflict with anyone's moral commitments. It will make them willing to do this if they give the reason that they have to find and comply with such principles absolute or near-absolute weight. That being so, no one reasonable will be firmly committed to *any* moral conception; her commitment will depend on how other reasonable individuals' moral views turn out to be. But they will be all in the same position. Therefore without coordination no moral conception will meet the ✓ hybrid contractualist test. Yet the coordinating mechanisms which could bring about that some moral conception meets the test seem to involve an inappropriate arbitrariness.

The arbitrariness puzzle is all the more puzzling in that the hybrid view which produces it seems, as I pointed out in Chapter 2, to capture something that modal contractualists wish to claim for their theory. This is the idea that the fact of reasonable individuals' motivation to find principles upon which all could reasonably be expected to agree makes a difference to the content of what all *could* reasonably be expected to agree upon. When the Individualist Restriction-respecting reasons that there are in support of some principle or moral conception are based on the relative levels of well-being, or the non-arbitrariness of distribution of goods, for example, it seems that being motivated to find and comply with principles based on those reasons is simply a matter of sufficient clear-sightedness and willingness; the reasons themselves are independent of whether anyone *is* this clear-sighted or willing. So the modal contractualists' claim is not justified. But hybrid theorists' identification of actual non-acceptance as grounding a reason for the rejection of some principle or moral conception changes this. Being motivated to find principles and moral conceptions that are best-supported by the reasons that there are *changes* the reasons that there are because it subjects one of the grounds of reasonable rejection to possible change itself. And this captures very well what

modal contractualists presumably have in mind when they claim that a reasonable individual is prepared to modify her claims in the light of the reason that she has to find and comply with principles that everyone could reasonably be expected to accept.

The arbitrariness puzzle arises, evidently, when reasonable individuals give such weight to the reason that they have to find and comply with principles that everyone could reasonably be expected to accept that they are ready to alter any of their other moral commitments in any way at all in order to conform to that reason. A solution, then, must plausibly recast the interaction between a reasonable person's commitment to conform with that reason and her other moral commitments.

There are three desiderata. First—for it to be a solution to the puzzle in the first place—the result must be such that mere coordination is not normally sufficient for the hybrid contractualist test to produce a determinate result: a just moral conception must be determined at least in part by the content of individuals' moral commitments considered apart from their commitment to conform with the reason that they have to find and comply with principles that everyone could reasonably be

expected to accept. Indeed, the possibility must be left open that at least in principle no moral conception would be just in some cases because no moral conception could secure the actual acceptance of all reasonable people. Second, recasting the interaction between a reasonable person's moral commitments must involve giving sufficient weight to her commitment to the modal contractualist reason for her to be counted by a hybrid theorist as reasonable in the first place.

Someone who acknowledges that reason but who is not ready to modify claims deriving from her other commitments *at all* surely cannot be counted reasonable. If she were, the solution to the puzzle would come at the cost of undermining the hybrid view's appeal. Finally, the recasting must nevertheless give sufficient weight

to reasonable individuals' *other* moral commitments. If acknowledgment of the burdens of judgment forces us to accept that someone may reasonably believe that *extra ecclesiam nulla salus*, then our description of the interaction between that belief and her acceptance of the modal contractualist reason must offer a plausible rendering of the idea that she really *does* believe that *extra ecclesiam nulla salus*, that this is a real moral commitment for her rather than an arbitrary starting point in the modal contractualist negotiations. It is in satisfying this last desideratum that the description I have been giving to draw out the puzzle in the first place fails.

What we need, then, is to relax the extent to which giving appropriate weight to the reason that she has to find and comply with principles that everyone could reasonably be expected to accept makes a reasonable person ready to abandon her other moral commitments. A hybrid theorist must be modest in her claims about the weight of the modal contractualist reason: giving it absolute weight precludes the satisfaction of the first and third desiderata. But can she do this? Can she consistently affirm, that is, that we have reason to find and comply with principles that everyone could reasonably be expected to accept and at the same time that this does not have absolute priority over the reason that we (take ourselves to) have to comply with principles that others could *not* reasonably be expected to accept? I think that she can and that she should. The revised Rawlsian hybrid view is built in part on a recognition of the burdens of judgment, a recognition that each of us may reasonably take herself to be bound to comply with moral principles that others may reasonably reject. In the context of that recognition, the force of the hybrid theorist's assertion that we all have reason to find and comply with principles or moral conceptions that everyone can be reasonably expected to accept surely cannot be that it would be unreasonable for us to be anything other than ready to drop every commitment that an appreciation of the burdens of judgment shows that we

may reasonably have. But nor can it be that there is a specific degree by which we must alter our views to bring them more into harmony with others'. It is perfectly plausible to suppose that two people who see the moral landscape initially in exactly the same way might, in recognition of the reason that they have to find and comply with principles that everyone can reasonably be expected to accept, modify their views in different ways, to different degrees, and yet neither be complying more or less than the other with the modal contractualist reason.

Rather, the assertion that we all have reason to find and comply with principles or moral conceptions that everyone can be reasonably expected to accept must, I think, be read as a plea for us to be *ready* to reconsider and perhaps modify—but not simply ignore—those commitments. Being reasonable in this sense is analogous to being reasonable in the realm of aesthetic interpretation, for example: it is being willing to reconsider what one sees with a view to appreciating (or trying to appreciate) what others see and perhaps adjusting one's own impressions as a result. But the result of our reconsidering our moral commitments, like the result of reconsidering our interpretations of music and paintings, may be that we do not alter them at all. A person may reasonably conclude that that she simply cannot change her position.

This offers a satisfying interpretation of Nagel's comment that the unanimity he is interested in is one "which could be achieved among persons in many respects as they are, provided they were also reasonable and committed within reason to modifying their claims".¹⁵³ The problem for Nagel is that since he has identified already the considerations which count in favour of proposed principles—relative well-being and ability to pursue one's interests—and can determine, from these, the

¹⁵³ EP, p. 33.

content of a just moral conception, the unanimity he is interested in is the unanimity of a society in which everyone sees those same considerations, and gives them the same weight, as Nagel himself does. The unanimity the hybrid theorist is interested in is, by contrast, not something that can be known in advance like this. For the extent to which someone who is ready to reconsider and perhaps modify her moral commitments in appreciation of the reason that she has to find and comply with principles that everyone can reasonably be expected to accept actually *does* modify those commitments in appreciation of that reason is not something we can know without finding out what she actually does. To repeat: there is no specific amount of modification which is the amount that is reasonable. So we really are interested in a unanimity of reasonable people, rather than in everyone's coming to see things as we do.

This also offers us a solution to the arbitrariness puzzle. The content of a hybrid view for a given society will not be open to arbitrary determination in the sense that factors besides those privileged by the view itself (reasonableness, the reasonably differing moral judgments of individuals) will determine it. A coordinating mechanism might be necessary to solve information problems, but it will not, now, also be the arbitrary determinant of which moral conception is just. The view works as follows. The individuals whose actual non-acceptance is grounds for reasonable rejection of a moral conception are those who are reasonable in the following sense/~~they~~ they are ready to reconsider and perhaps modify the way they see the moral landscape in the light of others' differing judgments about it and the reason that they take themselves to have to find and comply with principles that everyone could reasonably be expected to accept. ~~If~~ some moral conception cannot gain the actual acceptance of each person who is reasonable in this sense—if it is not such that after reconsidering and perhaps modifying their

1.1.1
only
this
will
can
have
examin

judgments about the moral landscape each person is willing actually to accept it—then it is not reasonable to expect each person to accept it. We have, however, no way of knowing in advance, even when we know what people's moral commitments are apart from their recognition of the modal contractualist reason, which moral conception—if any—will meet this test. For how far it is reasonable to modify one's moral commitments in recognition of the modal contractualist reason and others' reasonably differing commitments is not something that can be known without knowledge of how far a reasonable person actually does modify her moral commitments in recognition of these things.

It is this fact which explains how the hybrid theorist can sidestep the puzzle of scope reconciliation. The puzzle was that by restricting scope to those who take themselves to have reason to find and comply with a moral conception that everyone can reasonably be expected to accept the hybrid view forgoes an important part of the attraction of actual contractualism. For that restriction appears to be at least in tension with the underlying motivation for caring about actual acceptance in the first place: appreciation of the burdens of judgment. As I claimed in Chapter 1, it makes sense for actual contractualists to restrict scope in ways dictated by the motivation for their actual contractualism. But the further restriction to those who take themselves to be subject to the modal contractualist reason is not warranted by anything inherent in the hybrid view's *actual* contractualism. At best something seems to go missing; at worst the possibility is raised that the hybrid view is internally inconsistent.

But the solution that I offered to the arbitrariness puzzle shows us how we can get around the puzzle of scope reconciliation too. The crucial element in the solution to the arbitrariness puzzle is precisely the fact that acceptance of the modal contractualist reason does not determine any specific degree to which a reasonable

in 9th week

person may reasonably be expected to modify her other moral commitments in the light of the reason that she has to find and comply with principles that everyone can reasonably be expected to accept. [This means—I repeat—that for any given society the moral commitments of any person who fulfils only the conditions set by a burdens-of-judgment-based *actual* contractualist specification of scope could be the same as those that she would affirm if she fulfilled also the full set of conditions set by the hybrid view's specification of scope—even if as a matter of fact she does not fulfil the further condition of acceptance of the modal contractualist reason which those full conditions include.] Now note the following feature of the hybrid view that I did not dwell on when I introduced it in Chapter 2. According to that view, a moral conception is just if it is reasonable to expect each person actually to accept it. I have concentrated for the most part on one aspect of this formula: that it is reasonable to expect a *reasonable* person to accept a moral conception just if she *does* actually accept it. But a second aspect is that it is reasonable to expect an *unreasonable* person to accept a moral conception not if all *reasonable* people accept it, but rather if everyone accepts it *or would accept it if they were reasonable*. We know that we only have reason to care about individuals' actual acceptance insofar as they are reasonable. We don't, that is, have reason to care about their actual non-acceptance insofar as it is *unreasonable*. But we do have reason to care about what individuals could reasonably be expected to accept, which is what they would accept if they were reasonable; and as it now turns out this gives us reason to care about the *actual* acceptance of even those who aren't reasonable.¹⁵⁴ For what these people could

What?

¹⁵⁴ This brings out the way in which Rawls's modal contractualism runs counter to the common supposition that (for Rawls) "unreasonable citizens are rightfully excluded from the constituency of

reasonably be expected to accept—what they would accept if they were reasonable—*might* be just what they actually accept anyway.

To see this more clearly, compare again a more straightforward modal contractualism according to which the absolute level of well-being of the worst off person, say, is what grounds the reasonable rejection and acceptance of principles.

Even someone who is unreasonable in the sense that she fails to recognise the reason that she has to comply with principles which maximise the absolute level of well-being of the worst off person nevertheless does have reason to comply with such principles. She has reason to do what she would do if she recognised this reason. Now suppose that what she would do if she recognised this reason partially determined the content of what she would have reason to do even if she did not recognise it. Suppose, that is, that rather than entailing compliance with maximin principles, recognising the reason that we have to find and comply with principles that everyone can reasonably be expected to accept entailed compliance with principles that were partially determined by a combination of what everyone actually would accept independently of the modal contractualist reason and their acceptance of that reason itself. This, of course, is what the hybrid view says. Again we would have to ask what everyone—even unreasonable people—*would* accept insofar as they were reasonable. But the solution to the puzzle for hybrid theorists showed us that what unreasonable people would accept if and insofar as they were reasonable might be no different from what they actually do accept. So although unreasonable actual rejections, insofar as they are unreasonable, are to be excluded from the determination of just moral conceptions, that doesn't in fact give us reason to

public justification” (Jonathan Quong, “The Rights of Unreasonable Citizens”, p. 314; cf. Erin Kelly and Lionel McPherson, “On Tolerating the Unreasonable”.)

ignore, as a basis for reasonable rejection, the actual non-acceptance of unreasonable people. Their acceptance must be sought as well: only then can we be sure that a moral conception really does meet the hybrid theorist's contractualist test. So we avoid the worry that we are ignoring those whom our actual contractualist motivations give us reason not to ignore.

The heart of the hybrid theory and its contrast with straightforward modal contractualism remain intact in this picture. Hybrid theorists think that what is just is what everyone can reasonably be expected to accept. But they also have the actual contractualist intuition that a moral conception's conflict with a person's all-things-considered moral commitments is grounds for the conclusion that she cannot reasonably be expected to accept it. The difficulty has been that a hybrid theorist's commitment to the reason (the 'modal contractualist reason') that we all have to find and comply with principles that everyone can reasonably be expected to accept implies that not *everyone*'s non-acceptance of some moral conception on grounds of its conflict with her all-things-considered moral commitments is grounds for saying that she cannot reasonably be expected to accept it. So we have had to work out what to say about people who don't accept the modal contractualist reason; the worry was that our conclusions would show that the hybrid view is not, after all, actual contractualist in any way.

But the worry was assuaged. Identification of actual non-acceptance as the only grounds for reasonable rejection safeguards this view's actual contractualism. Requiring that it be that of reasonable people, or unreasonable people if they were reasonable, does not jeopardise it for two reasons. First, because, unlike on a straightforward modal contractualist view, no *other* grounds for the reasonable rejection of a proposed moral conception are specified—such as the relative levels of well-being of the least well-off, for example—the restriction to those who are

reasonable does not make the acceptance condition a mere proxy for such other grounds. And second, because the indeterminate nature of the difference that being reasonable makes to a person's all-things-considered moral commitments means that we cannot say of any unreasonable person that she would, if she were reasonable, accept a moral conception that she actually does not accept, no one is excluded from the scope of those whose acceptance is necessary for the justice of a proposed moral conception. The hybrid view really does combine the attractions of both modal and actual contractualism.

5. *The hybrid view and acceptance*

Before we turn to Chapter 4 and the task of connecting the hybrid framework with Rawls's elaboration of justice as fairness as a political conception, it remains for us to consider exactly how the hybrid view specifies the variable of acceptance (see Chapter 1, section 3). Evidently the salient fact about actual acceptance, for the burdens-of-judgment based revised Rawlsian hybrid view, is that it indicates congruence between an individual's moral outlook and the moral conception being proposed, since the moral conception being proposed is elaborated from basic elements shared by all. So the hybrid theorist's assignment of the variable of acceptance should reflect this. But traditional construals of actual acceptance—in terms of consent, for example—won't do that. Even once we have constrained consent so that it is free, informed, considered, and so on—as even traditional actual contractualists may plausibly be said or modified to do—there will remain the possibility that a person might consent to something which conflicts with her own moral views. If she were in a very weak bargaining position, for example, it might

make sense for her to consent (assuming that she understood consent to be necessary for the imposition of a moral conception) to a moral conception that treated her or others in a way inconsistent with her own moral views. But this would not be a case in which the Rawlsian hybrid view would say that the moral conception that she consented to had been justified to her.

So we say instead that the hybrid view's construal of what it means for someone to accept a moral conception should be something along the following lines: that she endorses a comprehensive moral conception which includes or supports the grounds employed in the elaboration of that moral conception. But two concerns that someone might have about this are as follows. First, she might suspect that this construal of acceptance undermines the actual contractualism of the theory that employs it, rather as construing acceptance as something like 'consent to the truth' would. Second, even if she conceded that this construal of actual acceptance was indeed actual contractualist, she might nevertheless deny that it could be the foundation of a plausible actual contractualist view. She might deny, that is, that there is any reason to appeal to actual acceptance in *this* sense (as opposed to informed consent, for example) to justify the imposition of a moral conception.

To some extent, I have already laid out the materials for replies to these two concerns in Chapter 1, section 3, and in section 3 of the present chapter respectively, but it is worth pulling them together here in order to make the replies clear. To begin with the first. As we saw in Chapter 1, actual contractualist theories are legitimacy views, which is to say that they make the justice of a moral conception conditional upon its reflecting facts about the particular, distinct individuals that the society in question is composed of. The fact about each individual that actual contractualists take to be important is the fact of their actual acceptance. They do this because they suppose that it is plausible to invoke the fact

of an individual's actual acceptance in justifying the imposition of the moral conception to her, and they seek to justify the imposition of the moral conception to each person individually.

Now, appealing to individuals' consent to the coercive imposition of a moral conception would have the obvious advantage, as a construal of actual acceptance, of drawing on our everyday practice of taking on obligations by consenting to them. But that doesn't show that consent in some form is the only plausible construal of actual acceptance in an actual contractualist theory. As I suggested, the essential point from a structural point of view is that what is invoked as justification to someone for the imposition of norms upon her is an exercise of her agency. Without that, actual contractualism ceases to be distinct either from legitimacy views more generally or else from regulative conception views. For in that case its justification takes the form of an appeal either simply to *who* the individual is or else to the way the (moral) world is. This is what would be the problem respectively with construing actual acceptance in terms of membership of such-and-such a family, for example, and with construing it in terms of consent to the truth. The essential point can be respected, however, without construing actual acceptance in terms of consent. And this is what construing actual acceptance in terms of endorsement of a set of moral judgments does. Appealing to a person's endorsement of these moral judgments constitutes an appeal to what can be conceived as an exercise of her moral-epistemic agency: she *endorses* them, she *judges* them to be true. (The idea of the burdens of judgment encourages this conception of us as moral-epistemic agents.)

The question then is: is this something that can be *plausibly* appealed to in order to justify to someone the imposition of a given moral conception? Part of the worry here is assuaged simply by pointing out that a person's endorsement of

principles and values is, because of the normative nature of these things, endorsement of their implementation. If someone endorses a set of principles and values, with their normative implications which include, sometimes, the legitimacy of coercion in accordance with other principles in the set, then she just does endorse the normative implications. This is simply what is involved in endorsing the principles and values.¹⁵⁵

But a more challenging worry is that a person's endorsement of a set of principles and values won't stand as grounds for the justification of the imposition of a moral conception that can be elaborated from them in the way that her endorsement of the moral conception *itself* would, because she doesn't actually endorse the moral conception itself even if she *would* were she to think about it carefully, say. This is more challenging because it calls into question, after all, the actual contractualist nature of the hybrid view, as follows. If it's not the case that a person (call her S) *actually* endorses, in some sense, anything that can be elaborated in a way that follows from her conception of the principles and values that she actually endorses (call these PV), then we seem to be presented with the following dilemma. Either (a) we justify the imposition on S of a moral conception (call this MC) which is elaborated in a way that follows from her conception of PV by appeal to the fact of her endorsement of PV even though PV is not MC; or (b) we justify the imposition on S of MC by appeal to the fact that she would have to endorse MC if she considered the matter, given that MC follows from PV. The trouble with (a) is that it appears less plausible than (b); the trouble with (b) is that it's not obviously

¹⁵⁵ I do not assume, in saying this, that some form of motivational internalism is true. Endorsing a principle is endorsing its normative implications, but that does not entail being motivated to comply with it. It entails endorsement of the proposition that one ought to comply with it.

a justification which appeals to anything construable as actual acceptance—it doesn't seem, that is, to appeal to any actual exercise of S's agency. For S may never have considered the matter. It would make our theory too demanding to suppose that justice cannot be achieved until we bring each person to consider whether they endorse, as part of their moral outlook, the moral conception that we want to impose.

However, there is in fact no dilemma here. In appealing to the fact that S would have to endorse MC if she considered the matter, given that MC follows from PV, we *are* appealing to the fact of her actual acceptance—of PV. Given that S might reasonably *not* accept PV, our justification of the imposition of MC must appeal to her actual acceptance of PV in order to work. (This was the point that followed from the burdens of judgment: recall the example of the buyers for the art museum.) As I emphasised above, the essential feature of actual contractualism is that justification of the imposition of a moral conception on someone appeals to an exercise of her agency. That is what determines whether or not it is actual contractualist. Whether the view is *plausible* or not is a different question, and that depends on how plausible the particular appeal that we make to the particular exercise of a person's agency is. My claim is that it is plausible to appeal to the actual exercise of S's agency that is involved in her endorsement of PV, with its internal connection to MC, to justify the imposition of MC on S. This is not a far-fetched claim. In a different context, Ronald Dworkin claims that

[i]f a doctor finds a man unconscious and bleeding, for example, it might be important for him to ask whether the man would consent to a transfusion if he were conscious. If there is every reason to think that he would, that fact is important in justifying the transfusion if the patient later...condemns the doctor

for having proceeded... [T]he patient's hypothetical agreement shows that his will was inclined toward the decision at the time and in the circumstances that the decision was taken. He has lost nothing by not being consulted at the appropriate time, because he would have consented if he had been.¹⁵⁶

Now Dworkin is right about the plausibility of the justification, but mistaken to suppose that the key point is the patient's hypothetical agreement. Who knows what he *would* have agreed to had he been conscious? Perhaps he is a highly irrational man. Perhaps he would have taken a violent dislike to the doctor. Perhaps the bleeding would have undermined his ability to think straight. What's really underlying Dworkin's reasoning is the idea that the patient's *actual acceptance* of a certain outlook ("his will was inclined toward the decision") can plausibly be appealed to in justifying acting in accordance with principles that follow from that outlook. And this is precisely my claim also.

Someone might object to this as follows. S could, in response to the attempted justification of MC by appeal to her acceptance of PV, claim: "but I never accepted, and don't accept, MC!" Since this could be true, and the failure to accept MC not irrational (it's not irrational, for example, not to consider every implication of the principles and values that one affirms), surely it's the case that we can't simply go ahead and impose MC anyway, claiming to be justified on actual contractualist grounds?

This seems right. But it doesn't undermine the plausibility of an appeal to S's acceptance of PV if S *hasn't* claimed that she doesn't accept MC. (If she never considers the question, for example.) And it doesn't alter the fact that that appeal is

¹⁵⁶ Dworkin, "The Original Position", p. 19.

an appeal to S's *actual acceptance* and therefore part of an actual contractualist strategy of justification. It undermines the plausibility of the appeal to S's acceptance of PV only in the case where S *does* offer the non-acceptance claim.

In this latter case, meanwhile, it is not clear how we should proceed. It's not obvious that the actual contractualist justification that appeals to S's acceptance of PV has been shown to be altogether implausible. After all, S does, *ex hypothesi*, accept PV, and all other things equal it is plausible to appeal to her acceptance of PV in justifying the imposition of a moral conception which follows from it. The problem is that it's also the case that she rejects MC, and this seems to undermine the plausibility of the PV-based justification.

The norms of rationality apply to S, however, and so if MC really does follow from PV, the onus is upon S either to affirm MC in the light of her acceptance of PV or else to reconsider PV given her rejection of MC. The burdens of judgment show us how it can be reasonable to differ from others in one's judgments about the moral landscape, but they do not excuse inconsistency. There is, however, no clear reason to suppose that S ought to take one or the other of these two routes.

Perhaps MC includes principles which appear to her so plainly wrong that she thinks she must revise her acceptance of PV. On the other hand, she might, in appreciation of the fact that MC follows from PV, decide that her rejection of MC was a mistake. So there is no clear reason to suppose that we ought to privilege her endorsement of PV over her rejection of MC or vice-versa in thinking about the justification of the imposition of MC. We are in a position analogous to that of the children of a man who, days before his death but lucid and apparently rational, verbally disinherits one of them—for reasons which date back decades—but who discover, shortly after his death, a recently-made will which divides his estate equally among all of them. The legal question here may have a clear answer, but the moral

one surely does not: justification of either an equal division or the exclusion of the verbally disinherited child from any inheritance could with some plausibility be made by appeal to what the father actually accepted.

In the case of the dead man, nothing further can be done to determine what division of his estate would be justified. But in the case of someone who has expressly rejected a moral conception which is elaborated in a way that follows from principles and values that she endorses, we can wait to see whether she changes her mind about which principles and values she accepts or whether instead she changes her mind about the rejection. And since she has now raised the question of her own acceptance of the relevant moral conception and is faced with the fact of the inconsistency of her position, it seems to me plausible to suppose that if, after a reasonable period, she has effected no change in the principles and values that she endorses, we may take that to constitute tacit acceptance of the moral conception that she rejected. On the other hand, if she effects a change in the principles and values that she endorses, then our actual contractualist justification must plausibly appeal to her acceptance of the new principles and values if it is to succeed. In the meantime, we may say that the justice of the moral conception in question is indeterminate. But this indeterminacy, because short-lived, does not create a great problem for the hybrid view.

To repeat: none of this undermines the plausibility of the actual contractualist justification by appeal to S's acceptance of PV in the case where she does *not* consider the question of her actual acceptance of MC itself. This will be the case for many people (most people don't spend much time thinking about the justification of the political order at all). It does bring out the fact that the type of actual contractualist justification that underlies the hybrid view is not once-and-for-all in the way that a justification which appeals to a person's having consented is.

Consent cannot be undone, whereas a person can alter her moral commitments and so undermine a justification which appealed to them. But I think that this, if anything, is a virtue of the hybrid view. It interprets individualistic justification as an ongoing process, answerable to individuals as they are, not as they once were. Nevertheless, this is, because it appeals to an actual exercise of agency on the part of S, a genuinely actual contractualist interpretation of the individualistic ideal.

Chapter 4: Political Constructivism and the Hybrid View

1. Introduction	161
2. The problem of practical conflict	165
3. Shared premises and the priority of justice	168
4. Indoctrination	180
5. False beliefs and loose beliefs	182
6. Mistaken interests	187
7. Wrong background	195

1. *Introduction*

Rawls's political liberalism answers 'yes' to the following question: can we develop justice as fairness in such a way that when we come to ask the question of stability, we can answer that justice as fairness is stable for the right reasons? According to the hybrid contractualist reading of Rawls that I developed in Chapter 3, this answer is determined by whether or not justice as fairness can secure the actual acceptance, construed in terms of endorsement of the fundaments of the political conception of justice as fairness, of each member of society. Justice as fairness must pass an actual contractualist test.

In this chapter and the next, I draw on elements of Rawls's argument in *PL* to justify the expectation that justice as fairness, and more particularly its egalitarianism, can indeed pass the actual contractualist test that the Rawlsian hybrid framework sets. This will involve three distinct tasks. The first two, which I set about in this chapter, are as follows. First, I discuss the grounds supplied by

Rawls's 'political constructivist' strategy for supposing that justice as fairness can secure the support of the range of comprehensive doctrines likely to persist in a pluralist liberal society. In section 2 of Chapter 3 above I noted that Rawls's intention in discussing the overlapping consensus at length in *PL* is to illustrate, and not merely assert, that support for justice as fairness leaves room for a variety of accounts of the realms of non-political values. Here I survey, then, the various illustrations that he provides and connect them to the more ambitious actual contractualist test that I advocate. I suggest that political liberalism's shallow foundations and limited applicability clear some of the obstacles to an overlapping consensus of all comprehensive doctrines. But what fundamentally explains the expectation that justice as fairness passes the actual contractualist test is a variant of part of *Theory*'s 'congruence' argument together with illustrations of the way in which different comprehensive doctrines support the values of the political conception of justice as fairness and, finally, the idea that the values of the political realm have a certain inbuilt priority over other values.

Much of this explanation of how justice as fairness can pass the actual contractualist test relies upon the distinctive way in which Rawls approaches the question in the first place. When Rawls asks about the likelihood of individuals' acceptance—when he considers the likelihood of an overlapping consensus on the political conception of justice as fairness—he asks about the likelihood that individuals in a society *already* shaped by justice as fairness will accept it.¹⁵⁷ That is what it means to ask about the stability of justice as fairness rather than its acceptability in the abstract. This 'head-first' approach contrasts with many actual contractualist approaches, which, if they specify the motives of the contracting

¹⁵⁷ See *PL*, pp. 140-1.

parties at all, take them to be those of individuals shaped by and seeking escape from the stipulated moral backdrop.¹⁵⁸ It also contrasts, of course, with non-actual contractualist approaches to justice. Such approaches may concern themselves with stability, but obviously not stability in any sense which entails an actual contractualist test as opposed to, say, the assurance that conditions in a society governed on their conception would not deteriorate to a state of civil war.

(Consider, for example, the kind of interest a utilitarian might have in the stability of utilitarianism.)

This approach has some clear disadvantages. Foremost among them is that it invites the charge that its stability, if indeed it is stable, is of no particular value, since just about any moral conception might be stable once there was no one left in society who wasn't shaped by (or brainwashed into accepting) it. Dealing with this charge is the second of the two tasks which I address in this chapter. I consider four forms of the objection that the stability for the right reasons of justice as fairness is based on indoctrination. I argue in reply that justice as fairness is not reliant, first, upon anyone's having (by its standards) false beliefs or, second, upon anyone's failing to think through her beliefs. Nor does it rely, third, upon individuals' acting against their own interests. And, fourth, it does not rely upon a version of the actual contractualist test that is set against the wrong background. I conclude therefore that the indoctrination objection fails.

But the Rawlsian 'head-first' approach also has important advantages. So long as it *can* deal with the charge of indoctrination, it allows us to draw on plausible

¹⁵⁸ This is how I read Hobbes, Locke, and Nozick (see Chapter 1). Rawls would dispute my reading of Hobbes, whom he (tentatively) interprets as adopting the head-first approach. See *Lectures on the History of Political Philosophy*, pp. 30-4.

principles concerning individuals' moral development under the kinds of institutions that justice as fairness requires to help explain why individuals' actual acceptance, in the sense we seek, would be forthcoming. In other words: we can make use of much of the theorising about individuals' moral psychological development that is familiar from Part III of *Theory*. Moreover, whereas traditional actual contractualist theories can rely on at best only informed guesses about the motives of the contracting parties, shrouded in the mists of the past as these people, the moral backdrop, and their contract must be, Rawls's answer to the question what each person actually accepts is informed by our own experience of liberal institutions.

(Indeed, the problem with *Theory* as Rawls sees it is precisely that individuals who are brought up under liberal institutions cannot reasonably be expected to endorse the same comprehensive doctrine. This is a premise that we are prepared to grant because we have first-hand evidence for it as well as because of its moral-psychological plausibility.) When we ask Rawls's question of stability for the right reasons, then, we ask it with a reasonable hope of an informed answer. We are familiar with the range of comprehensive doctrines that individuals affirm in liberal societies and therefore well positioned to judge the likelihood of their hypothetically consenting to various proposed moral conceptions.¹⁵⁹

The challenge that faces us is therefore somewhat different from that which faces traditional actual contractualists—even dual/mixed condition actual contractualists whose theories display the kind of unity I talked about in Chapter 1. For although they too are interested in the moral commitments of each individual,

¹⁵⁹ Cf. Raz, “Facing Diversity: The Case of Epistemic Abstinence”, p. 6. Raz writes that the fact that Rawls's “firm starting point is the society of the here and now” is “one of the very attractive features of his position.”

and present their arguments as moral arguments which would motivate each individual to consent to the moral conception they advocate, the basis for their hypothesis of individual actual acceptance of those arguments is speculative at best. Individuals, though they may be alike in fundamental nature, differ widely in their conceptions of values, needs, and goods. We might hope for more from those with actual contractualist commitments than the mere assertion that everyone will or did, as a matter of fact, accept their regulative conception view in the first place. And this is what we find in Rawlsian political liberalism, which aims to show how the basis of justice as fairness is actually endorsed by people like us—people who affirm the various comprehensive doctrines that persist in liberal societies.

The third task which must be completed in order to justify the expectation that justice as fairness's egalitarianism can pass the actual contractualist test set by the Rawlsian hybrid contractualist framework is to show that it is plausible to attribute to each individual the specific interpretations of the elements out of which the political conception of justice as fairness is built that are necessary to sustain the difference principle. This I attempt to do in Chapter 5 below.

2. *The problem of practical conflict*

The idea of the overlapping consensus is that a variety of comprehensive doctrines endorse the political conception of justice as fairness—the moral conception whose essential feature is Rawls's familiar two principles of justice—“each from its own point of view”.¹⁶⁰ Each person who is part of the overlapping consensus affirms a comprehensive doctrine from which she derives reasons to support the political

¹⁶⁰ *PL*, p. 134.

conception of justice as fairness. For the Rawlsian hybrid contractualism I have been developing, this overlapping consensus must be a consensus of all members of society (see Chapter 3, section 4 above). That is what the thoroughgoing Rawlsian actual contractualist test requires.

Notice the difference between the idea that each person affirms moral grounds which provide support for the political conception and the idea that each person merely derives reasons *for her to support* the political conception from her comprehensive doctrine. This latter is not the foundation for an overlapping consensus. The idea of an overlapping consensus is that values from each different comprehensive doctrine provide direct support for the substance of the political conception of justice as fairness, not merely that each different comprehensive doctrine gives its adherent reason to support the political conception. For the latter is consistent with an understanding of the political conception as worth supporting, for example, until such time as it is possible to replace it with a conception of justice expressive of the values of the comprehensive doctrine in question. If everybody affirmed the political conception of justice as fairness only for this reason, what would obtain would be a mere *modus vivendi*: a state in which parties are “ready to pursue their goals at the expense of ... other[s], and should conditions change ... may do so”.¹⁶¹ A *modus vivendi* would not be a state in which the kind of actual acceptance that we are interested in is secure. The kind of actual acceptance that we are interested in requires an alignment of a person’s moral commitments and the proposed moral conception.¹⁶² The former should support the latter, or at least not conflict with it. In an overlapping consensus, the political conception and its

¹⁶¹ *PL*, p. 147.

¹⁶² See my discussion of ‘unified’ dual condition actual contractualism in Chapter 1 above.

account of justice and the political virtues are justified on the basis of moral grounds to individuals who actually accept those moral grounds.¹⁶³

If the overlapping consensus is secured, that will represent a triumph in the face of the following problem. A comprehensive doctrine implies a set of moral values, principles, and requirements that run across the board to cover most (if not all) actions, relationships, and situations.¹⁶⁴ That includes the political actions, relationships, and situations of the basic structure of society. The political conception of justice as fairness, meanwhile, entails at least some values, principles, and requirements of its own, although these are restricted in applicability to the basic structure. Given the wide range of comprehensive doctrines that we find in liberal societies such as that engendered by the institutions of justice as fairness itself, it is natural to suppose that the chief obstacle to an overlapping consensus will be conflict between the values, principles, and requirements of justice as fairness and those of at least some of those comprehensive doctrines. In particular, justice as fairness may require that the world be one way while individuals' comprehensive doctrines require that it be another. It will therefore fail the actual contractualist test. This is the problem of practical conflict.

We have seen (in Chapter 3, section 2) that the point of developing the political conception in accordance with the strategy of political constructivism is to address the problem of practical conflict as it arises with comprehensive conceptions of justice such as that of the Kantian Interpretation of justice as fairness from *Theory*. But political constructivism does not suffice on its own to show that the overlapping consensus can be secured. The overlapping consensus requires not just

¹⁶³ See *PL*, p. 147.

¹⁶⁴ See *PL*, pp. 11-5, and Chapter 3, section 2 above.

that practical conflict be less likely but that it be eliminated. This is why it is so important that the political conception has not only shallow but *shared* foundations.

Shallow foundations and limited applicability clear the way for an overlapping consensus by eliminating certain clear grounds for supposing that the overlapping consensus *couldn't* be secured.¹⁶⁵ If they are not stressed, we might suppose that what is being proposed is a comprehensive conception of justice. But they aren't sufficient on their own to show that the overlapping consensus *would* be secured under justice as fairness.

Yet for all we have seen so far, the assertion of the shared foundations of the political conception is no more than an assertion. Shallow foundations and limited applicability help to explain why we shouldn't contradict it outright. But they don't vindicate it. So what does?

3. Shared premises and the priority of justice

Two moves are necessary to solve the problem of practical conflict. The first is to attribute, plausibly, endorsement of the elements of the public political culture out of which justice as fairness in its political presentation is constructed to each member of society. We must argue not only that they are to be found in the public political culture but also that the public political culture—at least insofar as it includes these elements—is part of or implicitly supported by the normative views of the individuals whose acceptance we are interested in. Otherwise there is no reason to expect the shallowness of the foundations of the political conception to

¹⁶⁵ As does epistemic abstinence (see Chapter 3, section 2, note 103 above) if we are prepared to grant what I called the doctrine of the embeddedness of truth.

be sufficient for its non-conflict with individuals' comprehensive doctrines. The first move, then, is to show that the political conception has shared foundations. The second move is to show not only that these elements are shared in this way but that the conception of justice which is constructed out of them will have *priority* over other elements in individuals' moral views. Otherwise there is no reason to suppose that when what justice as fairness requires and what other values that a person affirms require conflict, as they may even if a person affirms the elements of the public political culture out of which the political conception is constructed, individuals will come down on the side of justice as fairness.

Rawls makes both of these moves. The first is partly constituted by his endorsement of a broader version of part of *Theory*'s argument that justice as fairness "generates its own support."¹⁶⁶ Rawls supposes that according to plausible principles of the psychology of moral development and of human nature, members of a Rawlsian society would come to endorse justice as fairness.¹⁶⁷ Scanlon calls the argument for this conclusion a "purely factual or causal argument".¹⁶⁸ It appeals to the consistency of justice as fairness with empirically plausible principles concerning the way in which children develop moral sentiments. In *PL*, Rawls advances an argument to the effect not only that individuals will have a sense of justice but that they will see themselves and society in the terms which underlie the political conception. He draws in his elaboration of this argument on the idea of a

¹⁶⁶ *Theory*, p. 177/154.

¹⁶⁷ See *Theory*, §§70-73 (principles of the psychology of moral development) and §74 (human nature) as well as §§75-76 (both). (This is only a rough classification of the subject matter of the relevant sections.)

¹⁶⁸ Scanlon, "Rawls on Justification", p. 158. Henceforth I shall call it the 'causal-factual argument'.

'conception-dependent desire', a desire to act on principles which are understood as part of a political ideal:¹⁶⁶

Plainly, for us the main case is the ideal of citizenship as characterized in justice as fairness. The structure and content of this conception of justice lay out how, by the use of the original position, the principles and standards of justice for society's basic institutions belong to and help to articulate the conception of reasonable and rational citizens as free and equal. Thus we have an ideal of citizens as such persons...not only are citizens normal and fully cooperating members of society, but further they want to be, and to be recognized as, such members...they want to realize in their person, and have it recognized that they realize, that ideal of citizens. [...]

Thus, the account of justice as fairness connects the desire to realize a political ideal of citizenship with citizens' two moral powers and their normal capacities, as these are educated to that ideal by the public culture and its historical traditions of interpretation.¹⁷⁰

So individuals come to want to be and be seen to be citizens according to the political conception's articulation of citizens because the institutions of justice as fairness engender that desire in them. As long as the moral psychology which this relies upon is plausible, this is a perfectly reasonable expectation.¹⁷¹ Notice that the

¹⁶⁶ Rawls takes this idea from Nagel. See his *The Possibility of Altruism*, pp. 27-76.

¹⁷⁰ *PL*, pp. 84-6.

¹⁷¹ Rawls stresses that the moral psychology of this argument is 'philosophical not psychological' (*PL*, pp. 86-8). But that does not mean that it is immune to considerations of empirical plausibility. Rawls's point is that we are not constrained to identify and use some empirical account of human

argument implies individual acceptance of the fundamental conceptions of person and society which the political conception articulates. Individuals would come to see themselves in a way which is necessarily articulated in terms of the political conception, which itself involves the object conception of society as a fair system of cooperation between citizens seen as free, equal, reasonable, and rational (see Chapter 3, section 2 above). Thus this argument involves the attribution to individuals of the ‘fundamental ideas’.

We need not rely only upon Rawls’s conjectures about the moral psychology of individuals in the Rawlsian society in order to grant the attribution of the fundamental ideas. Rawls often voices the thought that the fundamental ideas derive from “the main moral and philosophical conceptions of a constitutional democratic regime”.¹⁷² They are part of the nature of the enterprise of society itself *as it is understood in democratic cultures*.¹⁷³ That understanding of society brings in its train the fundamental ideas, including the appropriate conception of those involved and on what kinds of terms they should base their interactions.¹⁷⁴ We are not constrained, therefore, only to *imagine* the self-understanding that would be prevalent in a Rawlsian society. We can also look at our own more-or-less liberal,

nature as the basis of our view. We are constrained only to respect the *limits* set by human nature. As he writes: ‘Whether [this moral psychology] is correct for our purposes depends on whether we can learn and understand it, on whether we can apply and affirm its principles and ideals in political life, and on whether we find the political conception of justice to which it belongs acceptable on due reflection. Human nature and its natural psychology are permissive: they may limit the viable conceptions of persons and ideals of citizenship, and the moral psychologies that may support them, but do not dictate the ones we must adopt.’

¹⁷² *PL*, p. xli.

¹⁷³ See e.g. *PL*, p. 38, n. 41.

¹⁷⁴ See *PL*, p. 301.

more-or-less democratic surroundings. (As I pointed out above, this is one of the great advantages of the Rawlsian approach.) And the fact that there is a great deal of disagreement among us about exactly what contested ideals such as fairness require should not blind us to the plausibility of supposing that we do indeed share the view that society is, or ought to be, a fair cooperative system among equals conceived as free and capable of cooperation and the pursuit of conceptions of the good.

Because the fundamental ideas and their elaboration are components in a theory of justice with shallow foundations, attributing them to individuals who grow up under justice as fairness's institutions is compatible with attributing divergent fundamental moral commitments to those individuals as well. People may affirm the fundamental ideas but also affirm other moral values and principles which may at times come into conflict with those of the political conception. To show that such conflicts will not arise, or that if they do arise the values of the political win out, the second of the two moves required to solve the practical conflict objection must still be made. As Rawls says:

[W]e need the second, and complementary, part of the answer to how political liberalism is possible. This part says that the history of religion and philosophy shows that there are many reasonable ways in which the wider realm of values can be understood so as to be either congruent with, or supportive of, or else not in conflict with, the values appropriate to the special domain of the political as specified by a political conception of justice//History tells of a plurality of not

unreasonable comprehensive doctrines. This makes an overlapping consensus possible, thus reducing the conflict between the political and other values.¹⁷⁵

As I said above (section 2) the kind of support that the values of the political conception are supposed to get from the ‘wider realm of values’ as interpreted by different individuals in Rawlsian society is itself moral. The values are supposed to give us reason to endorse the political conception—not merely to uphold it until such time as we can impose a political order which is more suited to our fundamental commitments.¹⁷⁶ So the substance of the second move required to solve the practical conflict objection is to a degree a matter of examining the kinds of comprehensive doctrines likely to persist in a Rawlsian society and asking whether they include grounds for support and prioritisation of the values of the political conception. This task can be approached in two (complementary) ways, both of which can be found in *PL*. First, we can run through a range of likely comprehensive doctrines and ask whether they include values which would give someone who affirmed them grounds to affirm the values of the political conception. Second, we can consider the values of the political conception itself and ask whether they do not have their priority built into them in some sense.

ov wo
(“)

Rawls offers a schematic overview of the first of these approaches in a short survey of the comprehensive doctrines in his ‘model case’ of the overlapping consensus.¹⁷⁷ These are a “religious doctrine” with an “account of free faith”; “a comprehensive liberal moral doctrine such as those of Kant or Mill”; the

¹⁷⁵ *PL*

¹⁷⁶ See *PL*, pp. 147-8.

¹⁷⁷ The model case is introduced at *PL*, p. 145. The survey is at pp. 168-72.

utilitarianism of Bentham and Sidgwick; and an unsystematic view which includes both the freestanding political conception and “a large family of nonpolitical values” in which “each subpart...has its own account based on ideas drawn from within it, leaving all values to be balanced against one another, either in groups or singly, in particular kinds of cases.”¹⁷⁸ Rawls sets out a moral basis to be found within each doctrine apart from the religious doctrine for a person to endorse a liberal political conception, stressing that “[n]o one accepts the political conception driven by political compromise”.¹⁷⁹ Of religious doctrines he says only that “I shall suppose—perhaps too optimistically—that, except for certain kinds of fundamentalism, all the main historical religions admit of [an account of free faith]”.¹⁸⁰ This survey would plainly be inadequate as a full explanation of the basis of the overlapping consensus. Rawls intends it merely as an example, however, designed “to illustrate some of the possible relations between comprehensive views and a political conception.”¹⁸¹ So he does not take himself here to have explained exactly how *the* political conception—that is, justice as fairness in its political constructivist guise—is related to the various comprehensive doctrines in a Rawlsian society. But we see how diversity in comprehensive doctrines *needn’t* entail conflict with the political conception (or at least its principles). Nevertheless, more still must be said.

One reason that Rawls offers but that the view that I recommend *cannot* accept, of course, is that reasonable individuals accept liberal principles straightforwardly as a consequence of their acceptance of the burdens of judgment. I have argued that Rawls’s derivation of the basic liberties from the burdens of judgment is a mistake.

¹⁷⁸ *PL*, pp. 145 and 169.

¹⁷⁹ *PL*, p. 171.

¹⁸⁰ *PL*, p. 170.

¹⁸¹ *PL*, p. 170.

However, insofar as it is supposed to be a consideration in favour of the overlapping consensus and therefore *stability* for the right reasons, one assumption that this mistaken conclusion relies upon *is* something we can accept. That assumption is that there are many reasonable individuals: many, and many more than there are unreasonable individuals. For as we know, Rawls takes the fact of reasonable pluralism as a given.¹⁸² And he takes political liberalism to be addressed to reasonable individuals: an overlapping consensus on justice as fairness among *them* is its aim.¹⁸³ So if justice as fairness can secure stability for the right reasons, it is for the right reasons because of the overlapping consensus among reasonable people. But it is *stability* because, presumably, the presence of unreasonable individuals who are not covered by the consensus is nevertheless insufficient to bring about instability. Rawls writes of unreasonable views that “there will always be such views. But they may not be strong enough to undermine the substantive justice of the regime. That is the hope; there can be no guarantee.”¹⁸⁴ What justifies even the hope? Ultimately, it is Rawls’s view that individuals simply *are* (in the main) reasonable, that we can take this as basic. Endorsing Scanlon’s idea that we are simply *are* moved by the reason that we have to find and comply with principles that no one could reasonably reject, Rawls writes that “in setting out justice as fairness we rely on the kind of motivation Scanlon takes as basic.”¹⁸⁵ As we saw earlier, Rawls sees both aspects of his conception of reasonableness — acknowledgment of the burdens of judgment and the desire to find and comply

¹⁸² See e.g. *PL*, p. 36.

¹⁸³ Rawls explicitly says this at *PL*, p. 36.

¹⁸⁴ *PL*, p. 65.

¹⁸⁵ *PL*, p. 50, n. 2 (carried over from p. 49).

with principles that each person could reasonably be expected to accept—as “two related expressions of this desire”.¹⁸⁶

What this means from the point of view of our reconstruction of Rawls's egalitarianism on a hybrid contractualist foundation is this. If we can follow Rawls in taking the attribution of reasonableness to most individuals as a given, then we can take it as a given that most individuals are ready to reconsider and perhaps modify their fundamental moral commitments, the way they see the moral landscape, in the light of others' differing judgments about it. For this is, as we saw in Chapter 3, what the hybrid theorist's interpretation of Rawls's first aspect of the reasonable amounts to. This means that in general there will be a pressure towards consensus—individuals will in general be aware of the differing judgments and ready to reconsider their own in the light of this. Moreover, this pressure would be towards consensus on justice as fairness. It will tend to erode conflicting views. Each person will have, as per the causal-factual argument, at the very least a disposition to accept the fundamental elements out of which the political conception of justice as fairness is constructed. To the extent that her comprehensive view tends to conflict with these elements and with the views of others, she will be ready to reconsider those aspects of it which conflict. The desire to find a moral conception which does not conflict with anyone else's fundamental views will tend to bring her to accept what each of them has in common even as the “free exercise of human reason” tends to lead her to moral conclusions which are not in other respects the same as those affirmed by others.

The survey of the ways in which the political conception may be seen to be congruent with a range of comprehensive doctrines, the tendency towards

¹⁸⁶ *PL*, p. 49, n. 2.

convergence created by reasonableness, and the existence of the disposition towards the values of the object conception predicted by the causal-factual argument combine to allay, to an extent, the worry that not only the nonpolitical but also the *political* aspects of individuals' comprehensive doctrines will conflict with the political values of the political conception. But it is worth emphasising that this is merely a tendency that we can plausibly expect, given the attribution of reasonableness, rather than a demonstration that the overlapping consensus will obtain. None of it is supposed to constitute proof. Rawls supposes that given sufficient numbers of reasonable individuals, their acceptance of the consequences of the burdens of judgment should secure an overlapping consensus on liberal principles. The hybrid view I am endorsing offers no guarantees even when everyone is reasonable; and in any case as we saw in Chapter 3 it requires in practice the actual acceptance even of the unreasonable, since their views might be no different if they were reasonable and what we care about is what they could reasonably be expected to accept. Still, a tendency towards the overlapping consensus on justice as fairness offers hope, as does Rawls's schematic overview of the ways in which radically divergent comprehensive doctrines may include within them the means to support the fundamental ideas of the political conception.

The second approach to the question of the priority of justice strengthens this hope. What was required, recall, was that it be explained not only how individuals come to share the fundamental ideas of the political conception, but also why there should be any expectation that they will prioritise the values expressed thereby. What Rawls says about the priority of justice is as follows. To begin with, he states that the "values of the political are very great values and hence not easily

overridden".¹⁸⁷ But what justifies this claim? Further on, during his discussion of the objection that the political conception must be comprehensive in order to provide guidance in "the many conflicts of justice that arise in public life",¹⁸⁸ Rawls writes:

values that conflict with the political conception of justice and its sustaining virtues may be normally outweighed because they come into conflict with the very conditions that make fair social cooperation possible on a footing of mutual respect.¹⁸⁹

In the discussion immediately preceding this passage, Rawls appears to be arguing that because a Rawlsian distributive scheme enables everyone to pursue her conception of the good, it "removes from the political agenda the most divisive issues"—an argument which seems simply to ignore that there might be a conflict of the sort we are discussing in the first place.¹⁹⁰ It's not obvious how the quoted passage is supposed to follow from this. But it offers Rawls's best answer to the question in hand. The essential thought is that society, conceived as a co-operative system, is made possible only when the values of the political *do* normally outweigh other values that come into conflict with them. As Rawls says earlier in *PL*, "these

¹⁸⁷ *PL*, p. 139.

¹⁸⁸ *PL*, p. 154.

¹⁸⁹ *PL*, p. 157.

¹⁹⁰ See *PL*, pp. 154-7.

values govern the basic framework of social life—the very groundwork of our existence".¹⁹¹

What explains why the values articulated by the political conception take priority over other values advanced in reasonable individuals' comprehensive doctrines, then, is those individuals' recognition that society is the 'groundwork of their existence'—the structure which shapes their lives and enables them to do better for themselves than any might do on her own¹⁹²—and so that the values associated with its regulation are of paramount importance. They must be given priority because otherwise cooperative society is threatened. Since the values associated with its regulation are the values of justice (the virtue which regulates cooperation), and since individuals tend towards support for the political conception in the ways that the first move in reply to the practical conflict objection suggests, the values of the political conception will have priority over other values in individuals' comprehensive doctrines.

¹⁹¹ *PL*, p. 139. It is also worth emphasising once again the intuitive importance of justification of a system of coercion. See *PL*, p. 137. Rawls develops the idea of the basic structure (which is the realm of the political as Rawls understands it) as the ground of our existence as the people we are with the culture and conceptions of the good that we have in Lecture VII of *PL* ("The Basic Structure as Subject"). See especially pp. 269–71.

¹⁹² The first of these two features is one of the grounds Rawls gives for making the basic structure the primary subject of justice; the second is an idea that Rawls puts forward at times in *Theory* (for more on which see Chapter 5 below).

4. Indoctrination

All of this takes us only so far. The causal-factual argument explains why we should expect individuals to conceive of the enterprise of society in a certain way that is essential to the development of the political conception of justice as fairness. The review of the comprehensive doctrines in the 'model case' shows us broadly how we might find grounds for support for the political conception in individuals' deeper moral commitments. The assumption of reasonableness gives us hope that individuals' deeper moral commitments will tend not to stray too far from the affirmation of values which do support the political conception. And the idea that society is the groundwork of our existence suggests that individuals will prioritise the values of justice.

But two large difficulties remain. One of them is that even if we are prepared to accept the causal-factual argument's attribution of the object conception to each individual, such adjectives as 'fair', 'free', and 'equal' are open to interpretation which we might expect to differ radically from one comprehensive doctrine to the next. It needs to be argued that the understandings of these concepts employed in the construction of justice as fairness are plausibly attributable to members of the Rawlsian society. For if different individuals do not share the same understandings, the importance of the 'groundwork of their existence' will not in all cases entail the priority of the values of justice as fairness as opposed to the values associated with whatever they see as the appropriate elaboration of the object conception given their different understandings of the concepts it invokes. In Chapter 6 I shall make a case for the plausibility of attributing to individuals in Rawlsian society a Rawlsian understanding of the fairness that underpins justice as fairness's egalitarianism.

For the remainder of the present chapter I shall address the second large difficulty, which is as follows. As we have just seen, to a large extent the expectation that justice as fairness in its political constructivist presentation will be stable for the right reasons is founded upon the supposition that individuals will share the fundamental ideas out of which it is constructed. They will, that is, view society as a fair system of cooperation between individuals regarded as free, equal, reasonable, and rational. But the plausibility of this supposition is itself explained to a large extent by the causal-factual argument, which says that among other things individuals in the Rawlsian society will as a matter of fact come to conceive of society that way. Yet if the explanation of stability for the right reasons—of a lack of conflict between anyone's comprehensive doctrine and the values of justice as fairness—is ultimately that each person's upbringing tends to bring about congruence between comprehensive doctrine and the values of the political conception, we might worry that the stability, someone might object that essentially the Rawlsian claim is simply that we needn't worry about the conflicts in question because we can brainwash people into conceiving of society and the values of the political as we want them to. Its stability might be for the right reasons in one sense—it isn't reliant upon coercion which individuals couldn't reasonably be expected to accept—but that doesn't capture what seems to be important about 'the right reasons' in the first place.

I can see four different ways in which someone might press the charge that political liberalism, in the form that I have been arguing for, is reliant on indoctrination. I distinguish and describe these below. I shall argue that the charge is unfounded whichever version we adopt. The four versions are as follows:

1. Political liberalism is reliant upon individuals' having false beliefs (*the false beliefs objection*)
2. Political liberalism is reliant upon individuals' failure to think through their beliefs (*the loose beliefs objection*)
3. Political liberalism is reliant upon individuals' acting against their own interests (*the mistaken interests objection*)
4. Political liberalism is reliant upon individuals' having been brought up in conditions other than those under which they ought to have been brought up (*the wrong background objection*)

I shall address them in this order.

5. *False beliefs and loose beliefs*

The false beliefs objection is obviously a charge of indoctrination. The thought underlying one form of it would be as follows. Political liberalism, in order to meet the necessary condition of stability for the right reasons, requires an overlapping consensus on justice as fairness in its political presentation which in turn requires each individual to conceive of society as a fair system of cooperation between individuals conceived as free, equal, reasonable, and rational. It requires, that is, that each individual should have certain beliefs about the nature of society. These beliefs are false. Therefore political liberalism relies upon individuals' having false beliefs. Alternatively, if it is argued that reliance upon beliefs that happen to be false, as opposed to reliance upon the *inculcation* of such beliefs, is not reliance upon indoctrination, the objection can be extended as follows. If each individual does

conceive of society in the relevant way, it's because (as the causal-factual argument shows) the institutions of justice as fairness engender individual moral development which involves individuals' coming to conceive of it thus. So political liberalism relies upon the inculcation of certain beliefs about the nature of society. These beliefs are false. Therefore political liberalism relies upon the inculcation of false beliefs.

This is not a convincing objection. It is true that political liberalism relies upon the inculcation of the object conception of society. But it is not obviously true that the object conception is false, in the sense that it is the *wrong* conception of society. Moreover, if it *is* false, then this objection is not primarily an objection based upon the idea that political liberalism relies upon indoctrination. It is an objection to the content of justice as fairness itself as a regulative conception. If it were primarily an objection based upon the idea that political liberalism relies upon indoctrination, it would be best employed as a *modus tollens* argument against the regulative conception. That is, the objection would run: justice as fairness, as a regulative conception, can be stable and therefore acceptable only by inculcating certain beliefs about society; therefore it may inculcate those beliefs; but inculcating those beliefs constitutes indoctrination and therefore is not permissible; therefore justice as fairness cannot be acceptable. However, this objection is not best employed thus. It is best employed as a direct challenge to the content of justice as fairness: since justice as fairness is elaborated from fundamental assumptions about society as a fair system of cooperation between free, equal, reasonable, and rational individuals, if those assumptions are false, then we have reason to reject justice as fairness so elaborated.

If the object conception of society were false that would, of course, undermine any claim that a friend of Rawls might make to the effect that the stability of justice

as fairness may be adduced as supporting grounds for it. To that extent the charge that it is false works as a charge of indoctrination: the friend of Rawls adduces justice as fairness's stability in its favour and the objector replies that this stability relies upon indoctrination since justice as fairness is based on false assumptions. But the primary charge is that the assumptions that justice as fairness relies upon are false, not the charge of indoctrination. It's analogous to the charge that utilitarians rely upon indoctrination just because the stability of utilitarianism would depend upon individuals' accepting the theory itself, which is false. Contrast this with the charge that they rely upon indoctrination because they rely upon individuals' believing a moral theory that they *themselves* take to be false, in the sense highlighted by Williams's 'Government House utilitarianism' jibe.¹⁹³

The false beliefs objection might also be put as an analogue of that jibe, however. Here the thought would be not that the object conception is false, but that Rawlsians themselves take it to be false even as they rely on individuals' acceptance of it in what they see as a just society. But there is no reason to suppose that Rawlsians *do* take it to be false. First, we should remember that Rawlsians and we members of liberal Western societies alike are taken to affirm reasonable comprehensive doctrines of the sort that would be affirmed by some individuals in the Rawlsian society, and so we are supposed to be able to endorse the object conception and justice as fairness as consistent with our own views. When we advance the theory we are not taking a stance which we suppose no individual in the Rawlsian society itself could take.¹⁹⁴ Second, note that Rawls's head-first approach

¹⁹³ See Williams, "The Point of View of the Universe", p. 166.

¹⁹⁴ See *PL*, p. 70: "As for the point of view of you and me—the point of view of the full justification of justice as fairness in its own terms—this point of view we model by our description of the

can be seen as an aspect of a unified dualist theory in which a regulative conception is affirmed but also subjected to a legitimacy test. Justice as fairness is affirmed but subjected to the test of stability for the right reasons. The regulative conception component does not involve only the assertion without any justification of justice as fairness: it is not an arbitrary moral conception inculcated simply for the sake of unanimity. It involves arguing from premises that Rawls himself accepts—the first third of *Theory*. Rawls makes plain in his later work that he views *Theory*'s argument for justice as fairness as sound, quite apart from the question of its stability.¹⁹⁵ So to the extent that justice as fairness is elaborated from the object conception in the regulative conception component of the dualist view, it is clearly not the case that Rawls himself does not accept the beliefs that are inculcated in individuals under the institutions of justice as fairness. Third, consider the fact that Rawls endorses in both his early and his later work the 'full publicity condition', which

has to do with the full justification of the public conception of justice as it would be presented in its own terms. This justification includes everything that we would say—you and I—when we set up justice as fairness and reflect why we proceed in one way rather than another. At this level I suppose this full justification also to be publicly known, or better, at least to be publicly available...if citizens wish to [carry philosophical reflection about political life that far], the full justification is present in the public culture, reflected in its

thought and judgment of fully autonomous citizens in the well-ordered society of justice as fairness. For they can do anything we can do, for they are an ideal description of what a democratic society would be like should we fully honor our political conception."

¹⁹⁵ See e.g. *PL*, pp. xvii-xviii and 6-7.

system of law and political institutions and in the main historical traditions of their interpretation.¹⁹⁶

It is clear that Rawls sees this as refuting objections of indoctrination in the sense currently under review:

When a political conception of justice satisfies [full publicity] condition, and basic social arrangements and individual actions are fully justifiable, citizens can give reasons for their beliefs and conduct before one another confident that this avowed reckoning itself will strengthen and not weaken public understanding. The political order does not, it seems, depend on historically accidental or established delusions, or other mistaken beliefs resting on the deceptive appearances of institutions that mislead us as to how they work.¹⁹⁷

This also answers the loose beliefs objection—that political liberalism is reliant upon individuals' failure to think through their beliefs. This can be seen as a charge of indoctrination because it implies that if individuals *did* think through their beliefs they would come to conclusions which would undermine the stability for the right reasons of justice as fairness. So political liberalism is reliant upon their not coming to those conclusions. If it prevents them doing so, it can be argued that it relies upon indoctrination. Even if it doesn't actually prevent them doing so but merely relies upon their not doing so, that seems objectionable in a similar way.

¹⁹⁶ *PL*, p. 67. It is clear that Rawls endorses full publicity (which is more extensive than the condition of publicity on the principles chosen in the original position) in *Theory* as well. See pp. 514-6/451-2.

¹⁹⁷ *PL*, p. 68.

What Rawls says about the public availability of the full justification of justice as fairness demonstrates that at least Rawls himself does not suppose that individuals' thinking through that full justification—as elaborated from the fundamental ideas—would imperil stability for the right reasons. So if the failure that the objector is referring to is the failure to reflect upon the full justification of justice as fairness, then certainly Rawls does not take himself to be relying upon any such thing. Moreover, none of the arguments I have been offering in this chapter to suppose that justice as fairness could be the focus of the overlapping consensus rely upon any failure on the part of individuals to reflect upon the compatibility of the fundamental elements with their comprehensive doctrines.

6. *Mistaken interests*

The mistaken interests objection is that political liberalism relies upon individuals' acting against their own interests. If justice as fairness's stability involves the inculcation in individuals of the object conception and the corresponding sense of justice, and this amounts to the inculcation of a motivation to act against their own interests, then what is at issue is plainly indoctrination. But even if the charge is not of inculcation of but merely reliance upon motivations contrary to individuals' interests, there is cause for concern. It would be good to show how political liberalism avoids it.

The worry here is what the congruence argument in *Theory* is designed to address. The congruence argument claims that an individual's sense of justice is congruent with her good, as understood on the thin theory of the good, in the sense that the thin theory supplies grounds for individuals to affirm and behave in

accordance with their sense of justice. (The thin theory is the theory of individuals' rational advantage drawn upon in the specification of justice as fairness. It is 'thin' in that it does not take into account the full specification of individuals' good that can only be offered once we know the content of justice.) Part of the argument offers reasons to deny that circumstances provoking conflict between the sense of justice and individuals' good will arise. But another part of it relies upon Kantian considerations which are supposed to show that the sense of justice is itself an essential part of individuals' good. Briefly, the idea is that individuals' true nature is expressed only in acting from the sense of justice, and it establishes the very strong conclusion that individuals would be irrational not to act in accordance with their sense of justice.¹⁹⁸ Because it relies on Kantian considerations, it cannot simply be re-invoked here: the political conception cannot draw upon elements of comprehensive doctrines.

We might try to answer the mistaken interests objection by running the same kind of argument but with an understanding of individuals' good which is given by a theory which does not include the controversial Kantian implications. So the challenge is to show first that we can (so to speak) de-Kant the thin theory (and also derive it without reference to any other controversial doctrine) and second that the congruence argument can successfully employ the de-Kanted version.

Now, the political presentation of justice as fairness is supposed to use a conception of individuals' good which is at least no more controversial than the fundamental elements of the political conception. Rawls argues that from the conception of the person implicit in the object conception of society, together with "various common sense psychological facts about human needs, their phases of

¹⁹⁸ See the references in note 105 above.

development, and so on”,¹⁹⁹ we can derive a ‘political idea’ of individuals’ good which “respect[s] the limits of, and serve[s] a role within, the political conception of justice.”²⁰⁰ This idea, as it turns out, is substantially very similar to the thin theory of the good advanced in *Theory*,²⁰¹ and grounds much the same list of primary goods as the interpersonal metric appropriate for justice as fairness. It could therefore be drawn upon in a version of the congruence argument.

How are we supposed to be able to derive the political idea of individuals’ good from these materials? The use of primary goods as the basis of interpersonal comparisons is said in *Theory* to be accounted for partly by the idea of ‘goodness as rationality’ and partly by “general facts about human wants and abilities, their characteristic phases and requirements of nurture, the Aristotelian Principle, and the necessities of social interdependence.”²⁰² They are the goods that “it is rational to want...whatever else is wanted”.²⁰³ According to ‘goodness as rationality’, a ‘human good’ is something which “has the properties that it is rational for someone with a rational plan of life to want”.²⁰⁴ A rationally planned life is a life that is “consistent with...principles of rational choice when these are applied to all the relevant features of [the person’s] situation and...is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that is,

¹⁹⁹ *PL*, p. 181, n.8.

²⁰⁰ *PL*, p. 176.

²⁰¹ See *PL*, pp. 176-7, n. 3, in which Rawls claims that “nearly all the structure and substantive content of justice as fairness [as presented in *Theory*] (including goodness as rationality) goes over unchanged into that conception as a political one”.

²⁰² *Theory*, p. 434/381.

²⁰³ *Theory*, p. 433/380.

²⁰⁴ *Theory*, p. 399/351.

with full awareness of the relevant facts and after a careful consideration of the consequences.”²⁰⁵ Choosing with full deliberative rationality is “careful reflection in which the agent review[s], in the light of all the relevant facts, what it would be like to carry out [various available] plans and thereby ascertain[s] the course of action that would best realize his more fundamental desires.”²⁰⁶ A good, then, is something which contributes to the satisfaction of the careful, considered, well-informed plan of someone who is trying to realize her most fundamental desires. This much, though fairly complex, seems uncontroversial enough to be part of the fundamental political conception of the person as rational (indeed Rawls says as much at *PL*, pp. 176-8). In *PL* Rawls also endorses basing the thin theory on the general facts about human wants and abilities and so on, adding only that we are also to be guided by the political conception of the person and emphatically not one (such as that of the Kantian interpretation) belonging to any single comprehensive doctrine. He writes that “to identify the primary goods we look to social background conditions and general all-purpose means normally needed for developing and exercising the two moral powers [for a sense of justice and for a conception of the good] and for effectively pursuing conceptions of the good with widely different contents”²⁰⁷ as well as “the framework of goodness as rationality and the basic facts of social life and the conditions of human growth and nurture”.²⁰⁸

²⁰⁵ *Theory*, p. 408/359.

²⁰⁶ *Theory*, p. 417/366.

²⁰⁷ *PL*, pp. 75-6. For the two moral powers see p. 19.

²⁰⁸ *PL*, p. 178.

Ultimately this involves “identifying a partial similarity in the structure of citizens’ permissible [comprehensive] conceptions of the good.”²⁰⁹ ‘Permissible conceptions of the good’ here means “those that respect the principles of justice.”²¹⁰ That makes it seem that we are to ask about the pursuit of which comprehensive doctrines would be permitted by justice as fairness and then find out which goods it would be rational to want regardless of which particular doctrine among these one affirmed. But that is surely a mistake. For as Rawls notes in *Theory*, “[t]he constraints of the principles of justice cannot be used to draw up the list of primary goods that serves as part of the description of the initial situation. The reason is, of course, that this list is one of the premises from which the choice of principles of right is derived. To cite these principles in explaining the list would be a circular argument.”²¹¹

We can, however, give a non-circular sense to Rawls’s restriction to permissible comprehensive conceptions. Consider how the structure of political liberalism suggests that the comprehensive doctrines that we are to draw on in formulating the list of primary goods should be identified. If the actual contractualist justification of the principles of justice arises from the fact that justice as fairness is the focus of an overlapping consensus (each person hypothetically consents to the political conception), then the basis of the use of the list of primary goods derives ultimately from the fact that this list is endorsed by all members of the consensus. But if the

²⁰⁹ *PL*, p. 180. Rawls often plays down this fact, suggesting for example that in formulating the list of primary goods “[w]e leave aside comprehensive doctrines that now exist, or that have existed, or that might exist” (p. 40). But identifying partial structural *similarities* must surely involve considering comprehensive doctrines as they now or might exist. What else are we to consider?

²¹⁰ See *PL*, p. 193.

²¹¹ *Theory*, p. 434/381. ✓

overlapping consensus is secured—and for the purposes of questions of indoctrination, that is the situation that we are asking about—then it is true that justice as fairness receives support from within each of the reasonable comprehensive doctrines in question; and in turn they each conform to the principles of justice. So the comprehensive doctrines among which we should look for partial structural similarities in drawing up the list of primary goods are members of the set of permissible conceptions of the good. Moreover, since stability for the right reasons is stability for many generations,²¹² the range of reasonable comprehensive conceptions likely to flourish, and which should therefore be included in the survey since they will be included in the overlapping consensus, will tend towards the full range of views in the set of permissible (but still plausible) comprehensive conceptions. Thus Rawls's restriction to permissible conceptions, though at first sight problematic, turns out to have an innocuous interpretation.²¹³

²¹² See e.g. *PL*, p. 3.

²¹³ This interpretation accords well with the parallel that Rawls draws between Scanlon's 'conventionalist' interpretation of his concept of urgency (of concerns) and the conception of primary goods as citizens' needs (*PL*, p. 188, n. 19). The conventionalist interpretation, based on usefulness in circumstances of pluralism, contrasts with a 'naturalist' interpretation which sees urgency as "the objective truth about which interests are more important and which are less so" (Scanlon, "Preference and Urgency", p. 668). If we begin at the outset with a survey of *actual* conceptions of the good in Rawlsian society we come closer to the conventionalist approach, I submit, than if we begin with a survey of *possible permissible* conceptions (though in fact as it turns out these may be close to coextensive). For the latter survey must be premised on the idea that primary goods are to be conceived not as *actual* citizens' needs but as *possible* citizens' needs. Since none of these need ever actually exist in order to do justifying work if they are included in the survey, it is less plausible to suppose that we are adopting an interpretation of citizens' needs based on its usefulness under the (actual) pluralistic circumstances that face us.

The thin theory and thus the primary goods can therefore be non-circularly and uncontroversially derived.

The thin theory gives us some reasons to suppose that the sense of justice is not contrary to individuals' good. First, as in the original congruence argument, we can point out that a Rawlsian society provides even its worst off members with "adequate all-purpose means" to develop and exercise the two moral powers—something which is a good for individuals who are conceived as essentially having an interest in developing and exercising those powers.²¹⁴ Second, we can argue that a Rawlsian society provides the social bases of self-respect, which also feature among the primary goods. Third, again following the original congruence argument, we can argue that participation in a well-ordered society of justice as fairness is experienced as a good.²¹⁵ This might depend on the Aristotelian principle, as it does in *Theory*; but in *PL* Rawls appears simply to draw on an empirical knowledge of democratic societies. We can also draw on Rawls's comments about special psychologies in *Theory*.²¹⁶ In other words, given much the same theory of individuals' good we can invoke much the same set of considerations in favour of the view that a sense of justice as fairness, and the Rawlsian society in which individuals are brought up to have that sense, will not conflict with individuals' good on that theory.

What we cannot invoke, however, is the most powerful of the arguments that Rawls gives in *Theory* to suppose that there will be no such conflict. That is the argument which invokes the Kantian idea that acting on the sense of justice is an

²¹⁴ See *PL*, pp. 202–3, and *Theory*, p. 169/147.

²¹⁵ See *PL*, pp. 204–5.

²¹⁶ See *Theory*, p. 541/474.

expression of individuals' fundamental nature. Without it, the considerations in favour of congruence that I have just listed may seem rather weak. In particular, they may seem merely to flag that justice as fairness isn't internally inconsistent, in the sense that the conception of individuals' good it employs at the beginning isn't undermined at the end. In *Theory* this is precisely what the congruence argument is supposed to do. But with that argument we are asking about a conception of individuals' good which was comprehensive. The question is: is the fundamental conception of the person which we have used in setting up justice as fairness, and which we assume to be true and comprehensive, supported or undermined by the society which we envisage under the institutions of justice as fairness? When we no longer take the fundamental conception of the person which we employ in setting up justice as fairness to be comprehensive, when we take it to be less than the whole story about individuals' good, finding out that it is not undermined by the society we envisage seems to leave important questions unanswered. We want to say: perhaps indeed the good on *this*, the political conception of individuals' good, is secured. But what about the good on *individuals' own* conceptions of the good—which conceptions the conception of the person employed in setting up justice as fairness no longer implies are simply mistaken?

This is a reasonable question, and it suggests that the mistaken interests objection is not fully answered by the appeal to congruence. However, now that we have distinguished the worry from that which is answered by the appeal to congruence, it should be clear that we have already answered it in dealing with the question how the various reasonable comprehensive doctrines in a Rawlsian society are meant to support the political conception of justice as fairness—how, that is, the overlapping consensus is supposed to be secured. For the answer to that—which appealed to the causal-factual argument, empirical facts about the doctrines likely to

flourish under justice as fairness, and the importance of the values of the political—is also an answer to the question: is justice as fairness consistent with individuals' good as they themselves conceive it? Their reasonable comprehensive doctrines are the source of our understanding of their conception of their own good; and these, it is argued, also contain within them the grounds of each individual's endorsement and prioritisation of the values of justice as fairness. If they don't, of course, justice as fairness isn't achieved anyway (so can't be said to be reliant on anyone's acting against their own good); and if they do, it evidently isn't secured through their acting against their own conceptions of the good. To the extent that we remain worried about indoctrination after hearing this reply, I suggest that that is because of the force of the fourth and final version of the charge of indoctrination: the wrong background objection. To that I now turn.

7. *Wrong background*

According to the wrong background objection, political liberalism is reliant upon individuals' having been brought up in conditions other than those under which they ought to have been brought up. If you remain unconvinced that political liberalism is not reliant upon indoctrination after reading my responses to the first three versions of the indoctrination charge, I think that this must be ultimately what worries you. The thought underlying this objection is that if justice as fairness is stable for the right reasons—that is, if it can be the focus of the overlapping consensus in the way that we predict—that is only because its institutions shape individuals under them so that they do not affirm any of a range of reasonable views

which involve rejecting the priority of the values of justice as fairness. So, for example, we can easily imagine someone who fits the following criteria:

- (1) she is prepared to modify her views in the light of others'
- (2) (even after such modification) she rejects the elaboration of fairness or the account of primary goods as citizens' needs which ultimately produce justice as fairness, or she doesn't see society as a fair system of cooperation

The actual contractualist test that Rawlsian hybrid contractualism sets requires this person's membership in the overlapping consensus.²¹⁷ Her endorsement is necessary for the stability for the right reasons of justice as fairness. But her comprehensive doctrine conflicts with the premises of justice as fairness, and so she could not be a member of the overlapping consensus. It would not be reasonable to expect her to accept justice as fairness. So stability for the right reasons would not be secured.

What the Rawlsian view I am arguing for says about this person is that she would not exist in the Rawlsian society. The account of primary goods as citizens' needs rests on structural similarities among *all* reasonable individuals' comprehensive conceptions of the good; *everyone* conceives of society as (properly) a fair system of cooperation; the elaboration of this idea of fairness is one which fits with *everyone*'s view. We predict all this in the light of the broad causal-factual

²¹⁷ We could in fact restrict the criteria to (2) alone and still have the problem because of the wide (unreasonable-including) specification of the variable of scope that I argued for in Chapter 3 above. It makes the objection clearer and stronger, however, to focus on the possibility of someone who is reasonable by the hybrid contractualist's own lights.

argument (see section 3 above), the tendencies created by the assumption of general reasonableness, and the priority built into values associated with society, the groundwork of our existence. Of course, it is possible that stability for the right reasons won't be secured, either because these predictions are unjustified or because though they are generally justified they may sometimes go unfulfilled. But that is not the objection. The objection is that when it *is* secured, and these predictions are correct, it is secured and they are correct because Rawlsian institutions, particularly in their inculcation of the object conception and sense of justice, make it so. The reason that the person does not exist is that Rawlsian institutions *prevent* anyone from becoming her, even though she cannot be condemned in the basic moral coin of reasonableness upon which political liberalism rests. But if she had not been prevented from existing, so to speak, then stability for the right reasons would not have been achievable. Stability for the right reasons is therefore nothing more than self-perpetuation through indoctrination.

I want to offer two replies to this objection. The first begins with the observation that no social world fails to shape the views of those who are brought up in it. That being so, it cannot be an objection in itself that individuals have been shaped by institutions. The objection has to be that individuals have been shaped by the institutions of the very moral conception acceptance of which by those same individuals is supposed to count as an argument for it. But what are the alternatives? Is there some social world that the objector can point to in which individuals *ought* to be brought up for their acceptance of justice as fairness to count as unindoctrinated? Two suggestions come to mind. First, a social world in which individuals are able carefully to consider the moral landscape and freely to endorse the moral conception that they conclude is best justified after such consideration. (This might be the suggestion of a Lockean.) In response to this suggestion we can

simply point out that the Rawlsian society, in which sufficient liberties and primary goods for the development and exercise of the two moral powers are guaranteed, and in which full publicity obtains so that people are brought up to understand the reasonableness of alternative ways of conceiving the moral landscape, is surely as close to the ideal formative social world as the Lockean objector could hope.²¹⁸

Indeed, it is one common criticism of the Lockean moral backdrop that for all its equality of basic liberties, Locke's own proviso on the acquisition of private property might permit²¹⁹ situations to arise in which consent to a moral conception would seem to be less free than the acceptance of justice as fairness by individuals in the Rawlsian society.²²⁰

²¹⁸ Notice that the causal-factual argument does not specify how children are to be educated in educational institutions. It discusses only the way in which individuals' moral sentiments are shaped by moral and political norms that are implemented around them. As regards education in schools, the Rawlsian view could easily support a normative analogue of the solution to the 'paradox of indoctrination' (education designed to bring it about that students form beliefs on rational grounds involves bringing students to form beliefs on non-rational grounds) proposed by James Garrison in "The Paradox of Indoctrination: A Solution". The idea would be that non-liberal worldviews are described as possible alternatives to the liberal worldview of justice as fairness in such a way as to create at least temporary suspension of belief in the liberal worldview. The causal-factual argument would not obviously lose plausibility in a society where this was the educational practice.

²¹⁹ Without anyone's simply giving away their legitimately-appropriated shares, a possibility that is generally (reasonably enough) ignored in discussions of what is permitted by various interpretations of Locke's proviso.

²²⁰ See e.g. Cohen, *Self-Ownership, Freedom, and Equality*, chapter 3. Cohen himself interprets Locke's proviso stringently enough that such situations probably wouldn't arise (see pp. 77-8). But he goes on to criticise Nozick on the basis of the relations that Nozick's interpretation of the same proviso permits between individuals—relations which are plausibly viewed as making some of those

The second suggestion is that individuals ought to be brought up in a social world in which there are no norms, or at least none belonging to a moral conception, so that individuals' acceptance of some moral conception is not shaped by the pre-existing moral conception. (This might be the suggestion of a Hobbesian.) But here it is important to distinguish between two senses in which a pre-existing moral conception may be said to shape the content of what individuals actually accept. In sections 4-6 of Chapter 1 I argued that a normatively empty moral backdrop such as Hobbes's would make an actual contractualist theory in one way attractive in that it would be, in a certain sense, thoroughgoing: every aspect of the moral conception ultimately advocated would be subject to the actual contractualist test of legitimacy. Moral backdrops which are not normatively empty—which include, that is, a moral conception already—do not hold this attraction because the (non-accepted) moral backdrop either prevents the possibility that certain moral conceptions may be accepted in a way that legitimises them (this happens when the moral backdrop includes inalienable or inviolable rights, for example) or else operates as a *moral* baseline which informs and morally endorses the bargaining positions of those individuals whose acceptance is necessary for the justice of any proposed alternative moral conception. So moral backdrops which are not normatively empty can be said to shape the content of what individuals actually accept in the sense that they impose norms which may at least inform if not specify those ultimately chosen.

But a different way in which a pre-existing moral conception may be said to shape the content of what individuals actually accept is via individuals' own

individuals less free than they would be under a Rawlsian moral conception (which indeed Cohen invokes as effectively more satisfactory in this respect at pp. 87-8).

commitments. Rather than shaping and endorsing their bargaining positions or restricting legitimising acceptance to only a certain range of moral conceptions, a moral conception might simply inform individuals' own moral thinking. To see the contrast between these two different ways of shaping the ultimately accepted moral conception, consider the Amish practice of *rumspringa* as it is popularly (mis)conceived. (Accuracy of attribution doesn't matter for my purposes.)

Rumspringa involves permitting teenagers to participate in wider society and practices without risking the severity of punishment that would be normally consequent upon such participation. They then have the choice either to return or to renounce the Amish life (with no sanction) as they see fit. These teenagers have been educated to see the world in moral terms that are set by Amish religious teachings, and thus they tend to see the outside world as morally inferior and choose to return to their home communities. Clearly they are shaped by Amish moral views. But it would be wrong to say that they make their choice from a bargaining position which is set by Amish values. Should they renounce their religion, it won't be the case that Amish conceptions of right and wrong continue to have the normative force they have for them as members of the Amish community. Nor is their choice restricted to a range of values determined by the Amish moral outlook. On the contrary, they can choose to embrace a completely different set of values with no normative shadow cast by Amish values.

By contrast, consider for example the wider laws which operate in the USA for the Amish and non-Amish alike. These do restrict the range of values that the Amish teenagers may embrace (or at any rate pursue), and they also affect the bargaining position that they hold in joining with others to pursue new moral lives. Perhaps, for example, a young Amish man decides that rather than return to the community he would like to found a new Amish community by converting non-

Amish people he finds in the outside world. But the extent to which he can pursue, as part of an Amish community, these values is very much influenced by the fairly liberal legal background against which he tries to found the new community. In the first place, a number of practices may be simply outlawed, so that even each individual's consent is insufficient to make them permissible. Furthermore, our young man's bargaining position is weak in the sense that others can easily refuse to accept Amish norms in favour of the norms already set by the liberal background: these background norms may allow them to take stands on the values that the community they found will have that a different set of background norms (more religious ones, for example) would not have facilitated.

A normatively (relatively) empty actual contractualist moral backdrop is analogous to there being, in our second example, no laws at all in the background. Evidently individuals will still have bargaining positions influenced by their situations prior to any agreement on the values of the new community. But these bargaining positions will not have the legal seal of approval that they would have if liberal laws were enforced. And of course there will be no practices which in advance we can say that the new community simply may not employ, since there will be no standard beyond that which they themselves agree. Such a normatively empty backdrop is perfectly compatible, however, with its being the case that the founders all have moral views influenced by their own moral upbringings. Our young man could still have been brought up to be Amish.

This being so, we must ask whether the suggestion about the appropriateness of individuals' acceptance not being shaped by the Rawlsian institutions under which they are brought up is a suggestion that the moral backdrop be normatively empty or a suggestion that individuals should not be shaped by those Rawlsian institutions. If it is the former, we can simply agree: it is perfectly compatible with a normatively

empty moral backdrop that individuals should have been brought up in a Rawlsian society. It is no part of the view I recommend that if justice as fairness fails to achieve stability for the right reasons, we must fall back on justice as fairness as the default moral conception anyway. If, on the other hand, it is the latter suggestion (that individuals should not be shaped by Rawlsian institutions), we can ask: how should they be shaped? It is utterly implausible to suppose that what we care about is the acceptance of individuals with no moral upbringing at all—if such a thing can be conceived. And it seems to me that the suggestion most likely to be acceptable to the objector is that individuals should be brought up to be, in some sense, autonomous in their moral judgment. But this takes us back to our reply to the ‘Lockean’ suggestion: the Rawlsian society, we can say, in which sufficient liberties and primary goods for the development and exercise of the two moral powers—which include the capacity to form and revise one’s judgment about the moral landscape—are guaranteed is surely as close to the objector’s ideal formative social world as we could hope.

I now turn to the second of my two replies to the indoctrination objection which claims that individuals brought up in a Rawlsian society are brought up under conditions other than those under which they ought to have been brought up, and that therefore the Rawlsian society is self-perpetuating only through indoctrination. The objection appealed, recall, to the possibility of a perfectly reasonable individual who nevertheless did not subscribe to the overlapping consensus. The thought was that therefore the prevention of such a person’s existence in a Rawlsian society must be the product of indoctrination on the part of the institutions of that society. The second reply to this is as follows: it is true that the Rawlsian society is inhospitable to certain views, many of which may be held by reasonable people. But that is not evidence of indoctrination. *No* moral conception would engender a society in

which all views, or even all reasonable views, flourished. This is part of the significance of Rawls's repeated reference to Isaiah Berlin's view that there is no social world without loss.²²¹ If the objector's point is simply that a Rawlsian society excludes some comprehensive doctrines which are not unreasonable, then either the indoctrination this implies is something that no moral conception can avoid or else something further must be said to vindicate employing this point as an objection against specifically the Rawlsian view.

One such further thing that the objector could say would be as follows. She could concede that justice as fairness might be stable for the right reasons, without indoctrination, once it has been implemented and perpetuated over generations. But wouldn't it have to rely upon indoctrination in the first place, in order to get started? For unless we have the good fortune to find ourselves confronted with a society in which everyone *already* accepts the fundamental ideas of the political conception, the only way to implement justice as fairness and have it meet the hybrid contractualist test of acceptance will be somehow to indoctrinate or fool people who wouldn't otherwise accept it. So either it relies upon such indoctrination somewhere along the line, or else it is condemned to irrelevance, since only societies that are already just by its lights stand any chance of achieving it, and there are no such societies.

Again, Rawls has a response to this objection. In the first place, he acknowledges that justice as fairness is indeed suitable only for certain societies at a

²²¹ Rawls makes reference to this idea at *PL*, pp. 5⁷, 18⁷ n. 2, 196⁷. See also *JAF*, pp. 36, n. 26, and 154. For Berlin's statement of the view, see "On the Pursuit of the Ideal"; and also "Two Concepts of Liberty", pp. 212-4.

certain stage in history.²²² A society without the democratic tradition which gives us the fundamental ideas from which justice as fairness is constructed cannot achieve it. The fact that the overlapping consensus could not be secured would be sufficient to show that a necessary condition of justice was not met. Now presumably a program of indoctrination could alter this situation so that the overlapping consensus could be achieved after all. But is that an objection to Rawlsian justice? Surely almost *any* theory of justice could see the conditions of its realisation achieved by a prior program of indoctrination. Suppose that merely reading a political philosopher's theory of justice to a person while she slept (i.e. so that no reasoning on her part was involved) were sufficient to bring it about that she would do everything she could to realise that political philosopher's conception of justice (consistent *with* that conception—so no murders, for example, if that account forbade murders for the sake of bringing about a just society). The number of philosophical accounts of justice which are inconsistent with their own realisation by means of the actions of individuals who had been read to in their sleep is surely zero. They might not endorse anyone's *reading to people in their sleep*, but given a society of people *who had been read to*, it's not clear that anything about those people would render their efforts to bring about justice on the relevant book's account delegitimising.

The possibility of its realisation by means of the actions of people who had been systematically indoctrinated in this way might constitute an objection to a theory of justice, I suggest, only when this possibility is *necessary* for the realisation of justice on that theory. I think that we are more inclined to criticise a theory of justice on these grounds if we cannot see how a just society on its conception could arise *except*

²²² See e.g. *J-4F*, p. 14.

by a pattern of systematic indoctrination.²²³ (Even then, if we are sufficiently convinced that the resultant society would be just, we might not count this an objection to the theory. But this seems less likely for an actual contractualist theory of the sort I am advocating.) So the dilemma facing the Rawlsian is this: either explain how justice as fairness could be realised without a pattern of systematic indoctrination, or face accusations of irrelevance.

In light of this objection we can see the significance of Rawls's story of the 'steps to overlapping consensus'.²²⁴ Rawls tells this story in response to the objection "that an overlapping consensus is utopian: that is, there are not sufficient political, social, or psychological forces...to bring about an overlapping consensus".²²⁵ The idea is that as a matter of empirical fact, it can be argued that there *are* sufficient forces present in a diverse range of societies for convergence on, first, a liberal constitutional consensus (in which liberal principles are accepted by all "as providing the only workable alternative to endless and destructive civil strife")²²⁶ and thence on the political conception of justice as fairness. What starts out as a liberal *modus vivendi*, accepted largely because of its necessity for peace, achieves a certain stability founded on three facts:²²⁷ first, it fixes the basic political rights and liberties so that the possibility of societal breakdown into destructive civil strife is pushed back over the horizon; second, its application and enforcement is relatively simple and can be specified in non-divisive terms and procedures, making it much

²²³ Of course it's *possible* that exactly the same set of actions that would be involved in a program of systematic indoctrination could come about by chance. I assume that we can safely ignore that here.

²²⁴ See *PL*, pp. 158-72.

²²⁵ *PL*, p. 158.

²²⁶ *PL*, p. 159.

²²⁷ See *PL*, pp. 161-3.

less open to suspicion as merely the cover for the advancement of factional aims; and finally, it tends to encourage citizens' reasonableness, which as we have seen creates a tendency towards consensus.²²⁸ Simple pluralism, Rawls claims, "moves towards reasonable pluralism."²²⁹ What then makes the constitutional consensus into the overlapping consensus on justice as fairness²³⁰ is the need for non-divisive discussion of policy with the aim of creating majorities as well as the need for some conception of the justification underlying the political system in place, in order to facilitate society-wide political self-understanding.²³¹

How convincing is this? It is true that liberal societies—at least societies with a liberal constitutional consensus of the kind that Rawls describes—*have* arisen around the world, and from diverse beginnings. Rawls's story about the steps to constitutional consensus (he tells a more detailed version in the introductions to *PL*) is at any rate not obviously implausible as an account of their origins. But even if it turns out to be false, it is also surely false that liberal constitutional consensuses have arisen as a result of the sort of systematic indoctrination which, if it were necessary for justice as fairness, would provide the basis for the objection. Of course, it might be claimed that the societies which take the first steps to

²²⁸ It is interesting that Rawls explicitly connects reasonableness with "a readiness to meet others halfway" in this section (*PL*, p. 163). This is not something that is obviously part of the Scanlonian understanding of reasonableness which Rawls elsewhere appears to endorse; but it fits very well with the characterisation of reasonableness that I have argued the hybrid contractualist must endorse to avoid the arbitrariness puzzle (see Chapter 3 above).

²²⁹ *PL*, p. 164.

²³⁰ Or some other moral conception among a class of liberal such conceptions: Rawls claims of justice as fairness only that it is central in this class.

²³¹ See *PL*, pp. 165-7. I take it that this last point is an acceptable gloss on Rawls's comments on pp. 166-7.

constitutional consensus are the way they are as a result of systematic indoctrination. So does justice as fairness rely upon *their* having indoctrinated their members? Not in the sense that our objector was worried about. The thought was that justice as fairness excluded through indoctrination a range of views not conducive to its own stability. Such views are *not* excluded by the moral conceptions prevalent in the diverse first-step-taking societies which have converged on the constitutional consensus. So the constitutional consensus, at any rate, cannot be accused of relying on their doing so. We are, then, halfway there.

What of the move from the constitutional consensus to the overlapping consensus? Does *that* require systematic indoctrination? Again, Rawls's story seems plausible enough, which suggests that it might not. But it is also worth highlighting his comments on the construction of justice as fairness and the role of political philosophy. At one point he writes:

No political conception of justice could have weight with us unless it helped to put in order our considered convictions of justice at all levels of generality, from the most general to the most particular. To help us do this is one role of the original position.

Political philosophy cannot coerce our considered convictions any more than the principles of logic can. If we feel coerced, it may be because, when we reflect on the matter at hand, values, principles, and standards are so formulated as ones we do, or should, accept. We can use the original position to further this recognition. Our feeling coerced is perhaps our being surprised at the

consequences of those principles and standards, at the implications of our free recognition.²³²

The dilemma that Rawls faced was this: either show how the overlapping consensus could come about without systematic indoctrination, or be condemned to irrelevance. I have suggested that the *constitutional* consensus certainly can (and has) come about without systematic indoctrination. But clearly there is no overlapping consensus on justice as fairness anywhere as yet. Now, however, we can turn to the argument of *Theory*, its development of the original position from assumptions that it is contended are fundamental elements in *our* democratic culture, part of the liberal constitutional consensus to which we subscribe, and consider it in a new light as an attempt to persuade us to adopt justice as fairness and thereby bring about the overlapping consensus. So far I have said much about the forces that tend to make justice as fairness, once achieved, stable for the right reasons. But I have not offered any articulation of the central idea: the political conception of justice as fairness. It is the putative fact that this is an elaboration of the fundamental ideas that all citizens share which plays the most important role in securing stability for the right reasons. Similarly, it is the putative fact that this is an elaboration of the fundamental ideas that *we* all share which will play the most important role in bringing about the overlapping consensus in the first place. If we are convinced by Rawls's argument for the principles of justice, and if we can convince others, so that justice as fairness can be implemented, this will not have

²³² *PL*, p. 45. Cf. p. 27: we employ the original position "to show how the idea of society as a fair system of cooperation can be unfolded so as to find principles specifying the basic rights and liberties and the forms of equality most appropriate to those cooperating, once they are regarded as citizens, as free and equal persons."

come about through systematic indoctrination, and the objection will be met. In Chapter 5 I try to meet it in this way insofar as it applies to justice as fairness's egalitarianism.

**PART III: HYBRID
CONTRACTUALISM
AND
EGALITARIANISM**

Chapter 5: Hybrid Contractualism and Egalitarianism

1. Introduction	211
2. The informal argument, stage 1	215
3. The informal argument, stage 2	223
4. The informal argument, stage 3	228
5. Reciprocity and impartiality	232
6. Conclusion	243

1. *Introduction*

An egalitarian conception of justice with the hybrid contractualist foundations that I have been describing in the preceding chapters stands or falls by the plausibility of its analysis of what it is that each person actually accepts. So far, we have seen that Rawls's political conception of justice as fairness is supposed to be developed from the object conception of society as a fair system of cooperation between citizens regarded as free, equal, reasonable, and rational. The object conception is something which each individual in a society of justice as fairness would share despite their disagreements about the correct comprehensive conception of the good. Yet surely, we might ask, these divergent comprehensive conceptions would tend to give competing meanings to the various concepts in the object conception? In particular, those hoping to vindicate Rawls's egalitarianism might suspect that the idea of a fair system of cooperation is open to too many divergent construals for there to be any consensus on the derivation of (any single set of) egalitarian principles from it.

My aim here is to argue that Rawls's distributive egalitarianism is derived from a construal of fairness which it is plausible to attribute to each person in a Rawlsian society. In the light of our conclusion in the preceding chapter that part of the argument for justice as fairness appeals also to the acceptability of its bases to all in liberal democratic pluralist societies such as our own, the aim is also to show that this construal of fairness is something that individuals in such societies do in fact endorse as part of their moral outlooks. I don't take this to be solely a matter of empirical surveys, however. Rather, it will involve arguing that the construal of fairness we are interested in is itself constructed out of uncontroversial components.

Success in this chapter, then, will be constituted by reasoning which plausibly takes us from ideas that it is plausible to suppose that we share—that is: you, me, those who share a liberal democratic culture, and Rawlsian citizens—to conclusions which support the egalitarianism of justice as fairness. (Rawls's difference principle is a highly egalitarian principle by most standards.)²³³ Rawls takes the argument from the original position to be largely independent of the argument of *PL*: the thought is that if we can connect its premises to the overlapping consensus, then

²³³ Nagel, for example, says it is 'strongly egalitarian' at "Equality", p. 109. (Cf. *Equality and Partiality*, p. 73.) Raz classes it as 'strictly egalitarian' at *The Morality of Freedom*, p. 232. Some philosophers think that it is not egalitarian *enough* (see e.g. G.A. Cohen, "Where the Action Is: On the Site of Distributive Justice") and others that it is not egalitarian *at all* (see e.g. Larry Temkin, "Equality, Priority, and the Levelling Down Objection", especially p. 134). To a large extent, however, these philosophers dispute the propriety of categorising the difference principle with principles which give the value of equality (as they understand it) a more dominant role rather than disputing that the difference principle produces outcomes which are highly egalitarian (Rawls emphasises that the inequalities permitted by the difference principle would not be great at e.g. *JAF*, p. 67). Cf. Véronique Munoz-Dardé's discussion of the difference between 'political' egalitarianism and philosophical egalitarianism in Chapter 2 of her *Bound Together: Claims of Others and the Needs of Selves*.

the fact that it is to be viewed as a deductive argument means that we have done enough to show that there is an overlapping consensus on its conclusions. I have suggested that there seems to be a gap between what the overlapping consensus is a consensus on (society as a fair system of cooperation) and the specific sense of fairness which grounds the egalitarianism of the difference principle. The argument of this chapter is intended to close that gap. But it is independent of the hybrid contractualist justificatory structures that I have been discussing hitherto in the same way that the original position itself is.

As a result, the argument will resemble other arguments for egalitarianism familiar from contemporary political philosophy in that it will be an attempt to persuade the reader that her own normative intuitions about fairness support egalitarian conclusions—that if she thought through those intuitions carefully and rationally, she would come to endorse the conclusions.²³⁴ These contrast with arguments which attempt to derive egalitarianism from rationality or natural facts, for example. Sometimes an appeal to intuition in moral argument can appear to be a disappointing dead end, especially when it seems but a short hop from intuition to conclusion. In the hybrid contractualist framework I have been elaborating, however, arguments which are based ultimately on such appeals are on a sure footing, for if we take pluralism and the burdens of judgment seriously we accept already that at the root of people's normative views are (and should be) intuitions about the moral landscape. That is the conception of moral reasoning that we build into our view.

²³⁴ Cf. for example Temkin, "Inequality: a Complex, Individualistic, and Comparative Notion", p. 334.

The dangers of this approach, from the hybrid contractualist point of view, are as follows. On the one side, there is the risk that the elaboration of fairness will be insufficient to explain the egalitarianism of the difference principle at all. It might be that minimal accounts of fairness (such as one, for example, that construes it as merely non-arbitrary application of principles) are uncontroversial in the sense that each person accepts that they should inform the construction of principles of justice, but that they are nevertheless insufficiently strong to ground the difference principle. On the other side, there is the risk that the elaboration of fairness will rely on normative assumptions which are too strong for it to be plausible to suppose that they are shared by everyone. The challenge is to steer a course between these dangers.

The familiar other arguments from intuitions about the badness of unfairness or the equal value of each person's utility might slot equally well into the hybrid contractualist framework, in the sense that they too aim at elaborating intuitions they take it to be plausible to attribute to us all. The plausibility of the overall view would depend in part on their plausibility as elaborations of our intuitions. But these arguments would not lead to the specifically Rawlsian egalitarian conclusions that I want to defend. In what follows I try to articulate the conception of fairness from which Rawls elaborates his egalitarianism and the fundamental assumptions from which that conception is elaborated. By starting from these assumptions, the argument in effect assumes that you accept them, and so that we can appeal to them as part of an actual contractualist justification of the imposition of the principles that are part of the argument's conclusions. Thus it attempts to vindicate Rawlsian egalitarianism as a view that we should endorse insofar as we are hybrid contractualists.

2. *The informal argument, stage 1: arbitrariness and the erosion of prior claims*

We know that the two principles are supposed to be derived, ultimately, from the object conception of society as a fair system of cooperation between citizens conceived as free and equal, as per the political presentation of justice as fairness. Evidently the conception of individuals as equal goes some way to explaining the egalitarian cast of the two principles. But it seems that there are plenty of other, less egalitarian ways in which a moral conception could express the equality of citizens. Advocates of inegalitarian libertarian theories, for example, do not suppose that on their conceptions individuals are seen as unequal.

What does the most substantial work in establishing Rawls's strong egalitarianism is the idea of fairness. This dictates the aspects of the original position—the veil of ignorance in particular—which determine the parties' choice of principles. So how does fairness connect up with equality? In *Theory*, Rawls appeared to be offering two arguments: that which proceeds from the original position, whose conditions are intuitively what we would accept as fair conditions for the hypothetical social contract, and what has since come to be known as 'the informal argument'. The original position argument, Rawls claimed, was the official argument; the informal argument was not, "strictly speaking, [an] argument for the [difference] principle".²³⁵ But in his later work Rawls ceases to distinguish so clearly between them. In particular, the idea that the principles of justice express a conception of 'reciprocity' which is articulated in terms drawn from the informal argument is explicitly invoked as an elaboration of the idea of fair terms of

²³⁵ *Theory*, p. 104, 89.

cooperation.²³⁶ So something akin to the informal argument should be seen as part of Rawls's official argument for the difference principle after all. Thus the original position, to the extent that it too is an elaboration from the idea of fair terms of cooperation, must also reflect this conception of reciprocity. Otherwise the original position cannot adequately elaborate the idea of fairness implicit in the idea of society as a fair system of cooperation (assuming that the difference principle is arrived at by only one such elaboration).

It is not altogether clear that the original position is based on the same elaboration of fairness that the idea of reciprocity expresses. The fairness of the original position seems to be, rather, the fairness of a bargain—indeed, Rawls says as much at the outset of *Theory*.²³⁷ The idea of a fair bargain seems different from the idea of reciprocity as Rawls elaborates that. But this is not important for my purposes here, since for the hybrid theory the aim is merely to elaborate from assumptions about fairness that are plausibly attributed to all to the egalitarianism of the difference principle. If this can be done without employing the original position as a device of representation,²³⁸ then there is no need to worry whether the original position also succeeds.

Let us work through the argument which takes us from a broad conception of society as a fair system of cooperation to specifically the difference principle, then. In *Theory*, Rawls writes:

²³⁶ See *PL*, p. 16, and *J-4F*, pp. 76-7 and 122-4.

²³⁷ See p. 12/11.

²³⁸ Rawls stresses that this is how we are to understand the original position: as a device of representation of premises that we accept. See *PL*, p. 27.

[A]lthough the difference principle is not the same as that of redress [the principle that undeserved inequalities call for redress], it does achieve some of the intent of the latter principle... The difference principle represents, in effect, an agreement to regard the distribution of natural talents as in some respects a common asset and to share in the greater social and economic benefits made possible by the complementarities of this distribution. Those who have been favored by nature, whoever they are, may gain from their good fortune only on terms that improve the situation of those who have lost out. The naturally advantaged are not to gain merely because they are more gifted, but only to cover the costs of training and education and for using their endowments in ways that help the less fortunate as well. No one deserves his greater natural capacity nor merits a more favorable starting place in society...the basic structure can be arranged so that these contingencies work for the good of the least fortunate. Thus we are led to the difference principle if we wish to set up the social system so that no one gains or loses from his arbitrary place in the distribution of natural assets or his initial position in society without giving or receiving compensating advantages in turn.²³⁹

But is the difference principle unfairly biased, he asks, towards the least favoured? Do the more favoured have a reasonable complaint?

Subject to the usual constraints (defined by the priority of the first principle and fair equality of opportunity), society could maximize the expectations of either [the best off or the worst off group] but not both, since we can maximize with

²³⁹ *Theory*, p. 87 in the revised edition.

respect to only one aim at a time. It seems clear that society should not do the best it can for those initially more advantaged; so if we reject the difference principle, we must prefer maximizing some weighted mean of the two expectations. But if we give any weight to the more fortunate, we are valuing for their own sake the gains to those already more favored by natural and social contingencies. No one had an antecedent claim to be benefited in this way, and so to maximize a weighted mean is, so to speak, to favor the more fortunate twice over. Thus the more advantaged, when they view the matter from a general perspective, recognize that the well-being of each depends on a scheme of social cooperation without which no one could have a satisfactory life; they recognize also that they can expect the willing cooperation of all only if the terms of the scheme are reasonable.²⁴⁰

Finally:

[T]he more advantaged representative man cannot say that he deserves and therefore has a right to a scheme of cooperation in which he is permitted to acquire benefits in ways that do not contribute to the welfare of others. There is no basis for making this claim.²⁴¹

²⁴⁰ *Theory*, p. 88 in the revised edition. The text of the original edition (p. 103) makes it less clear that the lack of antecedent claim to greater benefits for the more advantaged is what is supposed to drive the conclusion.

²⁴¹ *Theory*, p. 104 in the original edition; see also pp. 72-4 (62-4 in the revised edition). Although to an extent any inequalities which benefit everyone relative to the equal baseline do contribute to the welfare of the worst off group, it is nevertheless true that by comparison with those permitted by the

The first key claim of this argument, expressed in various forms in these three passages, is that the distribution of natural and social endowments is *arbitrary*.²⁴² What does this mean? By ‘natural and social endowments’ Rawls means the capacities and potential capacities, skills and potential skills, influences and possible influences, and upbringings the use of which tends to be remunerated in primary goods by others—i.e. which tend to have a high market value. By the *distribution* of natural and social endowments, Rawls means the way in which different people have different capacities etc. or have the same capacities etc. but in different degrees. And in claiming that this distribution is *arbitrary* Rawls means to highlight merely the lack of normative justification for it. Some might read the passages above as well as certain passages elsewhere as promoting the (stronger) view that the arbitrariness of the distribution of native and social endowments was morally objectionable. This would be a mistake.²⁴³ (I shall say more about this shortly.) Rawls is *not* implying that there is any distribution of native and social endowments which would be morally preferable.²⁴⁴ Nowhere does he suggest that something

difference principle, inequalities not permitted by the difference principle but still beneficial to all relative to equality include for the better off benefits which do not contribute to the welfare of others alongside those that do, in the sense that there are ways in which reduction in the total benefits for the better off could result in a higher total for the worst off.

²⁴² At *JAE*, p. 55 Rawls adds that the distribution of benefits and burdens arising from luck is also (of course) arbitrary.

²⁴³ See also for example *JAE*, p. 56; Rawls makes it clear that this would be the wrong reading at *JAE*, p. 76.

²⁴⁴ Cohen reads Rawls this way. See for example *Rescuing Justice and Equality*, p. 184. He does consider the view I am proposing here (pp.195-7), but rejects the consequent inferences to the difference principle as too egalitarian to be internally consistent. See below.

should be done about the distribution itself (such as attempting to equalise it, for example). As he writes,

[t]he natural distribution is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts. What is just and unjust is the way that institutions deal with these facts.²⁴⁵

The next step in the argument is to draw on the fact of the arbitrariness of the distribution of natural and social endowments to support the conclusion that no one *deserves* the set of natural and social endowments that she has. Rawls takes this to be obvious, I think, in its own right,²⁴⁶ but in any case it follows from the lack of any normative justification for the distribution of endowments. Nevertheless it is important because of the way it is appealed to in the next crucial step in the argument. This step amounts to the assertion that no one has a claim to advantages in terms of income and wealth/~~just~~ in virtue of her occupying a given place in the distribution of endowments. /

A great deal turns on this assertion, so it is worth focusing on it. It is not supposed to be the claim that people who are in some sense capable of commanding more primary goods than others outwith society (in some kind of state of nature) have no claim to those goods. For Rawls denies that *any* primary goods would be produced without social cooperation.²⁴⁷ Thus we are not supposed to see

²⁴⁵ *Theory*, p. 102 8⁷.

²⁴⁶ His defence of the claim in *JAF* rests on the rhetorical question: "Who would deny it?". See pp. 74-5.

²⁴⁷ He emphasises this only in *JAF* (see p. 61). But it is, as we shall see, hugely important. It undermines Parfit's suggestion that the unequal distribution of talents is equivalent to an unequal

the principles of justice derived from the original position as *redistributing* advantages that individuals would have had prior to or outside of the society regulated by those principles, whatever they turn out to require. The idea is not that no one has a claim to the fruits of her endowment (whether in cooperation or not), but that no one has a claim to any particular share of the fruits of *cooperation* just in virtue of her place in the distribution (among those cooperating) of endowments. Hence the distribution of natural endowments should not inform our choice of principles of distribution in one particular sense: we should not be concerned to respect anyone's claims on the primary goods produced through cooperation under such principles as we choose *when we make our choice of principles*. (Clearly once our choice of principles has been made we should be concerned to respect the claims on the primary goods produced through cooperation which are generated by those principles.)

The role of the preceding claim about desert here is as follows. At this point in the argument we have just moved from an assertion of the lack of any normative justification for the distribution of endowments to an assertion of the lack of any normative justification for those with endowments which have a high market value (in some particular market—no claims are being made about the absolute or cross-market values of endowments) to command a share of income and wealth proportionate to that value. It's not, however, clear at all that arbitrariness in the distribution of means of coming by something entails denial of one's claims to it. If I find a pound in the street because I happen to be the first person to pass by after

distribution of manna falling from the sky (see "Equality or Priority?", pp. 10-11). For the manna does not fall *only* when everyone engages in cooperation. Parfit implies that Rawls would agree with him that the inequalities in manna should be redressed. But this isn't clear at all. See below.

someone else has dropped it, the arbitrariness of the fact that I found it rather than anyone else does not ordinarily lead us to think that I have no more claim to it than anyone else. On the contrary, if restoration of lost property to its original owner proves impossible, we often suppose that it should be given to the person who found it. So it seems that the absence of normative justification for arbitrarily derived advantages nevertheless might *not* leave the field of justification completely open so that we can go on to select any principle of distribution that we like.

Instead, we might think that one's *having* the arbitrarily derived advantage gives principles which give one claim to it a *prima facie* head start over others.

Now the claim about desert operates, I think, not as a stage in the argument but as a rhetorical device intended to erode this kind of thinking. When it is observed that my finding the pound in the street was a piece of luck, it doesn't seem particularly strange to reply that that doesn't mean that I don't have a greater claim to it than you, who didn't find it. But when it is observed that I don't *deserve* my luck, this might be thought to undermine my greater claim to it compared to yours (all other things equal) more than does the observation that my being lucky is (obviously) arbitrary. The thought that people should not be permitted to hold on to undeserved advantages is more seductive than the thought that people should not be permitted to hold on to fortuitous advantages.

We shouldn't place much weight on the claim about desert. It is, of course, entailed by the arbitrariness of the distribution of natural talents. If arbitrary distributions are all that causes me to have something, then it's clearly not the case that I deserve that something. But pointing this out gives no greater force to the idea that I *shouldn't have* that something than does pointing out that the causes of my having it are arbitrary. In either case it is open to me to agree with the premise but reject the conclusion. As Allan Gibbard notes, "from the bare assumption that

fertility [in the example of differently fertile single-occupancy islands] is morally arbitrary, no obligation to share follows. The lucky ones could admit that their luck is morally arbitrary, and still ask ‘Why share?’²⁴⁸ What is needed is a *positive* reason to deny claims to the unequal shares to which the arbitrary distribution may give rise. ✓

3. *The informal argument, stage 2: equality and the significance of cooperation*

Some critics, faced with Rawls’s endorsement of equality as a ‘benchmark’ at around this stage of the argument, suppose that in order to supply a positive reason of the sort required Rawls *must* endorse the stronger version of the first claim in the argument after all—the claim, that is, that the arbitrariness of the distribution of endowments is in itself morally objectionable.²⁴⁹ For if *that* is the opening claim, then a distribution of income and wealth which reflected and therefore ‘rubber-stamped’ this morally objectionable distribution could plausibly be supposed to be morally objectionable itself. This would seem to open the way for us to ask which distribution would be morally preferable and then to answer: an equal distribution. But such critics fail to take into account the significance of Rawls’s supposition that no (significant amounts of) primary goods would be produced without cooperation. This is, as I said above, hugely important. It gives sense to such assertions as that

²⁴⁸ Gibbard, ‘Constructing Justice’, p. 269.

²⁴⁹ See for example Cohen, who says at p. 196 of *Rescuing Justice and Equality* that the weaker interpretation of the claim about moral arbitrariness with which the informal argument opens ‘does not...justify *beginning* with equality (or, indeed, with anything in particular)...it does not follow that inequality faces a case to answer.’

“some workable arrangement is a necessary condition of the good of all” and that “the well-being of each depends on a scheme of cooperation without which no one could have a satisfactory life.”²⁵

Why is this significant? The answer is that since cooperation is a necessary condition of the production of *any* (significant numbers of) primary goods, the distribution of the primary goods that a given cooperative scheme produces should incorporate acknowledgment of the contribution of all. And this gives us an answer to the question ‘why share?’ The answer is: ‘because these advantages are made possible only because of *everyone’s* contribution.’ That is not to say that no advantages could be made possible without a contribution from everyone. Rather, it is to point out that *these* advantages are made possible by the contribution of everyone. This is not an answer that can be straightforwardly given in the case of the found pound. It does not take social cooperation to turn my luck from a place in an arbitrary distribution into a financial advantage. But it does take social cooperation to turn my place in the arbitrary distribution of endowments into a share of primary goods.

This is the positive reason needed to deny claims to any unequal shares to which an arbitrary distribution of endowments among these contributors may give rise. Shares of any (significant) size depend upon everyone’s contribution; the question, now, is which particular set of shares is justified, given that fact together with the fact of the arbitrariness of the distribution of endowments. This opens up a normative space—not clearly opened up until we acknowledged the necessity of cooperation to the production of primary goods—which Rawls fills with a

²⁵ *Theory*, p. 103 ‘88. The revised edition omits the first version of the claim. See also pp. 126-8/108-9, *J4F*, p. 61, and below.

presumptive principle of equality. For since everyone's cooperation is a condition of anyone's having any primary goods, it will not do to take the default distribution to be merely as it happens to be under just any circumstances of cooperation. That distribution may not adequately reflect the fact that everyone's continued cooperation was a condition of anyone's having anything. An analogy may help to illustrate this. Suppose that I lift you into a tree to collect some apples. You collect the apples and jump down again. Then you give me a couple of apple cores that you don't want, and refuse to give me anything more even though you acknowledge that it was completely arbitrary that *you* were the one selected to pick the apples and *I* was the one selected to lift you into the tree to pick the apples and that therefore it's as a result of that arbitrariness that you *have* all the apples now. If you had merely *found* the apples on the ground, the arbitrariness of the fact that it was you and not I who found them would not imply that any principle of distribution of apples between us should be applied. But the arbitrariness of the fact that it was you rather than I who picked the apples opens up space for such a principle *given* that we cooperated to bring it about that there were apples available to either of us at all. And so you must justify your giving me only apple cores in the apple-picking example in a way that you needn't in the apple-finding example.

One might object at this point that even if cooperation removes any default claim to such advantages as one may receive under the circumstances of cooperation that happen to arise, the most natural choice of distributive principle to fill the space opened up by the acknowledgment of our dependence for any advantage upon the cooperation of all is one which awards each of us a share of the total product of cooperation in proportion to our contribution. What would be wrong with my giving you only cores in the apple-picking example would be, according to this thought, that your contribution to the enterprise was greater than

is reflected by the share that I give you of the overall proceeds. This principle, however, while admitting that not just any distribution will do in view of the necessity of cooperation, fails adequately to take into account the arbitrariness of the distribution of endowments. Even if we could isolate each person's contribution to the total product of cooperation (something that Rawls denies: see note 267 below), the point is that it is arbitrary that this is the contribution that each person is able to make. Importantly, this is not only because it is arbitrary that each of us cannot make more of a contribution to this type of cooperative enterprise, but also because it is arbitrary *that this type of cooperative enterprise is the one that is best for us all*. To see this, consider the apples again. Suppose that you are very light and I am very tall, and that we agree that my contribution to the apple-picking is smaller than yours, perhaps because we could have nearly as many apples if I were shorter (there are just a couple of branches that I couldn't lift you to) whereas we couldn't have even half so many if you were heavier (almost all the branches would break). Now it is still arbitrary, in the relevant sense, that you were the one selected to pick and I was the one selected to lift, because the distribution of lightness and tallness is no less arbitrary than the flipping of a coin. But it is *also* arbitrary that picking apples is the activity that produces the most food for us, and not, for example, tickling trout and then flipping them onto the river bank. My height enables me to reach right down under trout in a way that you cannot; your role in *this* cooperative venture would be to stop the trout from wriggling back into the river. We could have almost as many trout if you were heavier (you would be a little less nimble, so one or two more trout would escape you and wriggle back into the river) but we couldn't have even half as many if I were shorter (most of the trout hang around at my current arm's length below the surface of the water). If trout were more numerous, or apples trees less numerous, *my* contribution to our food production

would be the greater. Thus the contribution that we are able to make to a given scheme of cooperation is arbitrary in a second way. And so to each according to her contribution' is not a satisfactory choice of principle to fill the normative gap opened by the arbitrariness-of-endowments based denial of any claim to the shares achieved through cooperation as they happen to turn out (without including, that is, the results of any further principle of distribution).

The idea so far, then, is to undermine the thought that people have prior claims to any unequal share of the fruits of cooperation. We do this by (a) pointing out that the importance of each person's contribution derives from factors which are arbitrary: her own abilities and (more importantly) the suitedness of those abilities to the scheme of cooperation that society adopts; and (b) pointing out that *these* fruits of cooperation are made possible only through the cooperation of *these* contributors, i.e. the cooperating members of society. Even if one disputes the idea that one's talents are arbitrarily one's talents (perhaps because they are inextricably bound up with one's identity), it is hard to deny that it is only in the light of the suitedness to circumstance of the system which makes one's talents as socially useful as they are, together with others' willingness to cooperate in that system, that one could offer even the first premise in an argument for a greater share of the fruits of cooperation on grounds of productive talent. In the light of these facts, it seems that it would be

(1) *prima facie* unfair to set things up so as to reward people differentially for their role in the cooperative system. Moreover, it seems plausible to suppose that we are operating here with an uncontroversial conception of fairness. If society is a cooperative system, a process governing the way that individuals are brought together in order (among other things) to produce goods, then for it to be a *fair* cooperative system is for that process to be fair. But it would not be fair at all if it privileged some on the grounds of their fortuitous suitedness to this particular

system when others would have been better suited than them to different—but still cooperative, and productive—systems and, moreover, when their suitedness is in part the product of training and education which is itself the product of the cooperation of all. When all are contributing inputs to the cooperative system and none are entitled to their suitedness to that system, fairness requires equality.

4. *The informal argument. stage 3: from equality to the difference principle*

The *prima facie* claim of each to an equal share, ~~then~~, is the next stage in the argument, and it can then be finessed into the difference principle in recognition of the fact that some inequalities may lead to greater shares for everyone than anyone receives under an equal distribution.²⁵¹ The finessing works as follows. Since no one has a prior *prima facie* claim to any more than an equal share of the social and economic benefits that the cooperation of individuals with their various natural and social endowments brings with it, any inequalities in distribution must be of benefit to everybody, relative to the baseline of equality. Unequal distributions under which some have less than anyone would have under an equal distribution are therefore obviously impermissible. But there may be a number of unequal distributions which satisfy the condition that they should be of benefit to everyone relative to the

²⁵¹ Someone might object that the greater shares are objectionable all the same because of the various undesirable effects of the inequality they bring with them. However, Rawls denies that these undesirable effects would be likely given the degrees of inequality permitted by the difference principle (see *Theory* §§80-81 and the discussion of congruence in Chapter 4 above). To the extent that they would arise, they could be factored into the argument to limit inequalities in a way analogous to the way that Rawls supposes they would be factored into the argument from original position if necessary.

baseline of equal shares. How to choose among them? Rawls's thought is that we should simply do as well as possible for the group which does the least well out of the inequalities—that is, for the group which sees the smallest improvement in their lot from the baseline of equality, which is all anyone has a *prima facie* claim to in the first place. To set up the inequalities any other way—even if it remains the case that everyone benefits relative to the equal baseline—would amount to granting some better-off groups extra advantages (namely, the difference between what they would get under the difference principle and what they get under alternative unequal arrangements satisfying the all-must-benefit-relative-to-equality condition) just because of their suitedness to the system of cooperation that society ~~operates~~, a factor which we have acknowledged to be arbitrary. And it would be to penalise members of the worst-off group (again, by the difference between what they would get under the difference principle and what they get under the alternative unequal arrangements) for *their* suitedness to the system of cooperation—and this despite the fact that the cooperation of *everyone* is necessary for the production of these benefits. Recognition of the arbitrariness of individual suitedness to the system of cooperation together with the necessity of cooperation to the production of benefits at all, and the consequent claim that no one has any greater claim than anyone else to the benefits of cooperation, are inconsistent with a willingness then to tolerate inequalities, even those that benefit everyone relative to the equal baseline, which reward the naturally and socially well-suited to this type of cooperation and penalise the less well-suited in this way. As Rawls puts the point in *PL*:

Because they start from equal shares, those who benefit least (taking equal division as the benchmark) have, so to speak, a veto. And thus the parties arrive

at the difference principle... those who have gained more than others are to do so on terms that improve the situation of those who have gained less.²⁵²

The conclusion is that the expectations of the worst off group should be maximised. Anything else is “to favour the more fortunate twice over”.²⁵³ Thus we arrive at the difference principle without drawing on any controversial premises beyond those which took us to *prima facie* equality.

Let's highlight the point at which difference principle's strong egalitarianism enters into the argument. We find an obviously strong egalitarianism in the idea that the sum of primary goods to which no one has a prior entitlement deriving from natural productive talents should be divided equally among all. This creates the benchmark against which departures from equality are to be judged. But we also find what Raz calls 'weak' strict egalitarianism in the idea that the sum of primary goods to which no one has a prior entitlement—because the cooperation of all is necessary to the existence of that sum at all—deriving from natural endowments should be divided so that departures from the baseline of no entitlements at all for anyone are to the greatest benefit of the worst off.²⁵⁴

²⁵² *PL*, p. 282.

²⁵³ *Theory*, p. 88 in the revised version.

²⁵⁴ See *The Morality of Freedom*, Chapter 9. A strictly egalitarian principle has the form 'All Fs who do not have G have a right to G if some Fs have G'. A weak strictly egalitarian principle tolerates inequalities which are such that attempts to redistribute the extra benefit compared to an equal baseline cause it to disappear altogether (i.e. when we are dealing with 'lumpy' goods). Relative to a baseline of equal shares, the inequalities permitted by the difference principle appear to violate strict egalitarianism. But considered relative to a baseline of nothing for anyone, the unequal distribution permitted by the difference principle considered as a whole does not violate strict egalitarianism. The connection between the strict egalitarianism that the difference principle expresses and the

Notice that the informal argument does not involve the claim that each is independently entitled, for some reason, to a given stock of primary goods, a stock to which she would be entitled even if an equal division would not fully meet that entitlement, and that an equal division will as a matter of fact grant everyone that to which she is independently entitled. The informal argument does not employ what Raz calls a principle of entitlement.²⁵⁵ It employs an idea which is fully relational: *because* the cooperative contribution of all is necessary for anyone to have any primary goods, anyone's *having* primary goods indicates that everyone else should also have primary goods. The mere existence of inequalities in the distribution of primary goods, prior to any application of the difference principle, is sufficient to create the egalitarian entitlements of that principle.

But what makes room for the strong egalitarianism of the difference principle's treatment of the sum of primary goods produced through cooperation is Rawls's denial that the distribution of the benefits of cooperation needs no justification. The grounds of this denial are the claims that without the cooperation of all, these benefits would not be available for distribution at all and that one's degree of suitedness to the scheme of cooperation that we adopt is arbitrary. The questions are therefore, first, which scheme of cooperation to adopt and, second, how to distribute the benefits that arise from it. It is here that the strong egalitarianism of the difference principle makes its entrance: given that without the cooperation of all there would be no benefits, and given that it is arbitrary that for any given scheme of cooperation any given person is able to contribute as much as she does, we

reasoning of the informal argument may be stated as follows: if some Fs have G, all Fs have a right to G, because some Fs have G only as a result of the (only arbitrarily larger or smaller) cooperative contribution of *all* Fs.

²⁵⁵ *The Morality of Freedom*, p. 218.

should choose the scheme and distribution which are such that the person who benefits least is better off than the person who benefits least under any other scheme and distribution.

To summarise. No one should receive a greater share than others except where no one with a lesser share could do better than she does when the person with the greater share receives that share. This is because of the arbitrariness of the fact that the scheme of cooperation which offers the worst off as much as this scheme does requires the development of certain capacities which the better off have and not the development of certain other capacities which the worst off have, and which might lead them to be among the better off under some other scheme of cooperation.

The difference principle manifests a recognition of the sacrifice that the worst off make for the sake of a greater share for each person (where 'each person' does not, when unpacked, rigidly designate). Recall the apple-picking example: the difference principle as applied to the distribution of apples manifests recognition not only of the arbitrariness of my being tall and your being light, but of the sacrifice I make by not insisting that we go trout-tickling instead—that is, insisting on a cooperative scheme under which I would do better.

5. *Reciprocity and impartiality*

Rawls stresses, especially in his later work, that this conclusion expresses a conception of reciprocity.²⁵⁶ Reciprocity, he writes,

²⁵⁶ See especially *JAF*, pp. 64 and 76–7, as well as *Theory* pp. 102–6, 88–91 and *PL*, pp. 16–7.

is a relation between citizens expressed by principles of justice that regulate a social world in which everyone benefits judged with respect to an appropriate benchmark of equality defined with respect to that world.²⁵⁷

The point of reciprocity is that individuals “do not gain at one another’s expense since only reciprocal advantages are allowed.”²⁵⁸ Nobody is to benefit from cooperation unless their benefiting contributes to other cooperators’ benefiting too. Reciprocity is not mutual advantage, however, because to be motivated by reciprocity is not to be wholly self-interestedly motivated.²⁵⁹ It requires individuals’ support for a distribution which in some cases gives them less than other, feasible distributions of the same fruits of cooperation might. Rawls also claims that reciprocity is not impartiality because it draws nevertheless, as an ideal of impartiality does not, on the idea that cooperative society *is* to the advantage of all

²⁵⁷ *PL*, p. 17.

²⁵⁸ *Theory*, p. 104–89.

²⁵⁹ See Gibbard, “Constructing Justice”, pp. 266–71. Rawls’s own explanation of why reciprocity is not mutual advantage depends on the idea that if we were to transport people from some situation in which the distribution of holdings did not satisfy his own two principles into a situation in which they did, some would not gain (see *PL*, p. 17). But if we were to transport them from a situation in which no one had anything into the Rawlsian situation, everyone *would* gain. This merely serves to emphasise that we need some conception of prior holdings in order to articulate a conception of mutual advantage; it doesn’t show that reciprocity isn’t mutual advantage. Rawls does appear to offer, at times, such a conception: the ‘no agreement point’ of general egoism. See *Theory*, p. 147/127, and below.

(it is not altruistic).²⁶ This is important because it restricts the scope of the difference principle's egalitarianism to the cooperating parties. Non-cooperators are not owed any share of the fruits of cooperation on grounds of distributive justice.

I want now to consider an objection to the thought that the intuitions which support this idea of reciprocity are sufficient to sustain justice as fairness. The objection concludes that we must attribute stronger intuitions which support the ideal of impartiality instead. This is a serious objection, since the ideal of impartiality is much less plausibly attributed to each person in liberal democratic societies. Impartiality implies obligations as strong as those that we have to those with whom we cooperate even to those with whom we do not cooperate. It seems less likely, given our experience of life in liberal democratic societies, that each person accepts such obligations as part of her fundamental moral outlook than that she accepts ideal of reciprocity. The obligation itself is much stronger and the intuitions of fairness which would sustain it are also less plausible, as we shall see. So it would be good to refute the objection.

The objection starts with the idea that how well individuals would do in other cooperative groups must be factored into the question which distributive ideals would receive the support of each person. I have said that it is an assumption of the informal argument that no (significant) goods can be produced without cooperation; but that does not mean that no goods could be produced without cooperation among specifically *these* individuals, that is, the set of members of a Rawlsian or liberal democratic society about which we are asking whether justice as

²⁶ See *PL*, p. 17, n.18; and also *J4F*, p. ???. Rawls does not in these passages offer any argument for this; he merely endorses Gibbard's defence of him in "Constructing Justice". The explanation I offer is Gibbard's.

fairness meets the actual contractualist test (call that set R). So what happens when we factor the possibility of alternative cooperative groups, composed of subsets of R , into our thinking about distributive principles? Nozick (who suggests this line of thinking as part of an argument against the difference principle) thinks that support for the difference principle will be seriously undermined, since difference-principle-regulated cooperation among R might not be self-interestedly rational for all members of R compared with more limited cooperation regulated by principles to which all can agree.²⁶¹ We can also compare difference-principle-regulated more limited cooperation with that among all members of R . Some might very conceivably do better if they withdrew from R and cooperated among themselves.²⁶² (In the light of the trout-tickling example, it should be clear that these need not, however, be those whom we might designate ‘the talented’ in the cooperative scheme in which everyone participates.) At the very least, some might be able in light of this fact to negotiate better terms for themselves in the scheme of cooperation among all of R .

So although reciprocity is supported as a criterion for the distribution among members of R of the fruits of their cooperation by the intuitions about fairness that sustain the informal argument, it’s not clear that for some individuals those intuitions are sufficient to ground support for the cooperative scheme which includes all members of R as opposed to some other scheme(s) also regulated by the

²⁶¹ See *ASU*, pp. 192–7.

²⁶² Even taking into account the possibility that those withdrawn from might act aggressively towards them after they have withdrawn or as they are withdrawing. Gibbard seems to think that this possibility shows that no one could withdraw (see “Constructing Justice”, p. 272), though he ultimately adduces fairly similar considerations in support of his doubts about the plausibility of justice as fair reciprocity.

difference principle. A member of R can accept the reasoning in favour of the difference principle and yet reject, precisely because of the size of the shares the difference principle (which she accepts) allocates her, cooperative systems in which she does less well than she might—including the scheme of cooperation among all members R .

If this is right, then it one might suppose that Rawls must do something more than attribute the intuition of fairness to the members of a liberal democratic society in order to conclude that each member actually accepts the scheme of cooperation itself. For surely he needs to have something to say to prevent people 'shopping around' for cooperative schemes, and so to justify not only the relative shares that the difference principle allocates to cooperators in a given scheme, but also participation in that scheme and therefore those shares absolutely?

Two options might seem obvious. The first is to attribute a stronger intuition of fairness to cooperators. This stronger intuition would involve seeing the mere fact of the arbitrariness of the distribution of endowments and such goods as came to people other than by their own hand as sufficient on its own to ground a denial that *anyone's* shares of *any good* should be unequal—whether they were members of the same cooperative group or not. In that case cooperators would support moral conceptions which equalised the shares of individuals in different cooperative groups and outside any cooperative groups alike; and if these shares weren't equalised, they would see shopping around for cooperative schemes as trying to take advantage of one's place in the arbitrary distribution of endowments. Consequently either there would be no advantage to be gained by shopping around or else they would not shop around anyway. The second option is to attribute not a stronger fairness-based motivation but a kind of solidarity among the members of R which would lead them not to shop around even though they might do better by

participating in some other cooperative scheme. The thought here is that as before everyone hypothetically consents to reciprocity as the criterion for the distribution of benefits of any cooperation between them all, but now in addition we stipulate that no one is motivated to look to any other (reciprocity-governed) cooperative scheme in the hope of improving her lot.

But in both cases, according to the objector, the Rawlsian hybrid contractualist must presuppose motivations stronger yet than those that are presupposed according to my elaboration of the informal argument. Those presupposed in the second of the two options are, moreover, hard to square with the hybrid contractualist foundations I have been offering for justice as fairness. The identification of the actual contractualist test of legitimacy as a necessary condition of justice arises from the idea that it would not be reasonable to expect an individual to accept a moral conception which conflicts with her own normative commitments; the move to the political presentation of justice as fairness arises from the idea that a liberal society is a pluralistic one in which everyone will not rally round a single comprehensive set of such commitments. Indeed, members of a Rawlsian society may be very much opposed to one another's moral views in many respects. But we hope, with Rawls, that justice as fairness can, by proceeding from shallower, shared foundations, offer a basis for a public sense of mutual trust and respect. If, however, the account of justice as fairness's passing the actual contractualist test presupposes the kind of solidarity among members of society which would prevent them thinking about whether they might do better elsewhere, the appeal of justice as fairness as a *new* basis for solidarity is limited. So the second option is not attractive.

This leaves the first option. This was to make the difference principle's justification to all dependent upon their acceptance of the view that prior claims to

benefits howsoever arising are to be denied. This means that even those producing or finding benefits without cooperation would be denied any greater claim to them than anyone else. Rawls, of course, denies that any such production is possible. But it also means that those producing benefits through cooperation are denied any greater claim to them than anyone else, including those cooperating in other schemes. For whereas fairness as reciprocity appeals to non-altruistic motives—it involves giving fair (i.e. not arbitrary distribution-reflecting) return to each other when we have cooperated for our own benefit—the stronger idea of fairness that we are considering here involves giving fair return (i.e. not arbitrary distribution-reflecting) to *everyone*, regardless of whether they have been involved in a scheme for our benefit or not. Anything else, on this stronger idea of fairness, would be to take advantage of one's place in the arbitrary distribution of endowments.

This is the content of the ideal of impartiality. So the force of this objection to the informal argument is that Rawls cannot stop, in justifying the egalitarianism of the difference principle, at the attribution of the motivation of fairness required to elicit acceptance of the informal argument. The ideal of reciprocity which this motivation leads individuals to endorse is insufficient to prevent individuals from shopping around in the way described above. He must (so goes the argument) accept the ideal of impartiality and attribute the correspondingly stronger conception of fairness in the motivation of Rawlsian individuals.

But the objection fails. We are interested not in individuals' choices of cooperative scheme but in their acceptance of ideals for the regulation of cooperation in the case of a what Rawls calls a 'closed society', that is, one whose "members enter...only by birth and leave...only by death."²⁶³ That being so, we

²⁶³ *PL*, p. 12.

abstract from the possibility that anyone could ‘shop around’ for alternative schemes of cooperation: this is the only one possible (and the only alternative to be considered is one of no cooperation at all, which secures the idea that cooperation is for the good of all). The closed society assumption would itself be objectionable if it were part of the argument for individuals’ acceptance of justice as fairness that justice as fairness was to each person’s advantage compared to the alternatives. For while that is plainly true with respect to the alternative of no cooperation at all, it might well be, as Nozick suggests, untrue with respect to a benchmark of the shares that each individual might get in cooperative groups other than *R*. Some of the things that Rawls says in places might lead one to suppose that a benchmark of no cooperation is operating as the baseline in a bargaining process which produces the difference principle. For example, he says that “we can interpret general egoism as the no-agreement point. It is what the parties would be stuck with if they were unable to reach an understanding.”²⁶⁴ But these comments are offered as description of the situation facing the parties in the original position rather than a description of the situation for which justice as fairness as a whole is constructed.²⁶⁵ The function, in this latter context, of the idea that no goods at all would be produced absent cooperation is not to provide a benchmark for the purposes of ascertaining an agreement point in a bargaining process which each individual is

²⁶⁴ *Theory*, p. 136 / 117-8. See also p. 147 / 127; cf. pp. 15 / 13-4 and *PL*, pp. 278-9.

²⁶⁵ Notice that when Rawls describes the ‘circumstances of justice’ at *Theory*, pp. 126-8 / 109-10, the generality of the discussion suggests that Rawls views society *in general* as a cooperative venture for mutual advantage rather than that he takes any given instance of a society to be a venture for the mutual advantage of its members *as compared with other possible cooperative ventures*.

conceived as engaging in. It is not invoked as a point of comparison.²⁶⁶ Its function is, rather, is to highlight the necessity, in any given scheme of cooperation, of the contribution of all the cooperators to the specific fruits of that specific scheme of cooperation.

When we are asking about what would be a just distribution for a given scheme of cooperation, then, we are not interested in counterfactuals about how well individuals would do under the alternatives. So assuming away those counterfactual situations by considering only the case of a closed society does not alter the content of what it is that we suppose that each person would accept given intuitions of fairness. This brings out what is wrong with the objection. The objection takes it to be important to explain why individuals wouldn't 'shop around' for a better deal. But such shopping around would threaten justice as fairness only if the shoppers used the deals that they could get in other cooperative schemes to give themselves bargaining advantages in negotiating their shares under the Rawlsian scheme. However, their acceptance of reciprocity as the appropriate ideal for governing the relations of members of the Rawlsian scheme entails that they would not do this. So there is no need to explain why they would not shop around. It is true that we have nothing to say to stop individuals withdrawing from any given scheme of cooperation in order to join some other scheme in which they might fare better. But why should he have to have anything to say about this? Should he be distinguishing between individuals who seek to withdraw because they feel a sense of belonging in some other society and those who seek to withdraw because their talents might be developed differently and rewarded more? (Consider by analogy

²⁶⁶ I dispute, therefore, one common reading of Rawls: see e.g. G.A. Cohen, *Self-Ownership, Freedom, and Equality*, pp. 224-5, and Barbara Fried, "If You Don't Like It, Leave It", p. 40.

someone who withdraws from friendships because he doesn't feel able to carry the burdens that the duties of friendship impose. Should someone who is trying to describe the duties of friendship alter the content of those duties in light of the possibility of such a person?) Every other scheme will also be regulated by justice as fairness, given the motivations of those involved; and this is what we are interested in.

Conceptions of justice of the variety that Rawls aims to defend are conceptions of justice to be applied to the relations between the members of a single society. The actual contractualist challenge is to show that the members actually accept this conception of justice. The informal argument offers reasons to suppose that individuals motivated to cooperate on fair terms with those with whom they cooperate—as we have supposed that individuals in Rawlsian society, each sharing the object conception of society, would indeed be motivated—would accept, in the sense we are interested in, the difference principle as regulative of their cooperation. Stronger motivations would indeed be required if those individuals were also to accept that they should not consider their prospects in alternative schemes of cooperation. But justice as fairness does not seek to show this.

Could it be objected that refusing to offer an account explaining why individuals would not consider their prospects in alternative schemes leaves Rawls open to the possibility of instability? After all, if individuals *were* constantly considering their prospects in alternative schemes, it seems unlikely that a stable society could emerge. The answer to this objection is twofold. First, we can invoke once again the congruence argument and in particular the claim that individuals would have sufficient means to live decent lives and so, given fair cooperative terms which each accepts, would have no strong material incentive to seek alternative schemes. We can also note that individuals' conceptions of the good are in general not focused on

the need to maximise one's absolute stock of income and wealth. Second, and more prosaically, we can point to the costs of withdrawing cooperation. The possibility that individuals will withdraw cooperation in order to form or join cooperative schemes in which they might have better expectations in terms of primary goods is real, but unlikely to be realised to any great extent. This is simply because it is costly (not only in terms of primary goods) to leave one's society and costly to join and form one also. If there was injustice in these costs, this might be an objection in itself. But it is no part of the objector's case that there is such an injustice. And there is no reason for us to adopt the view that there is one either.²⁶⁷

²⁶⁷ One worry that the criterion of reciprocity gives rise to is the possibility that those who do not contribute to the social product, but whom we would ordinarily see as members of society, are not owed a difference-principle regulated share. For the necessity of *these* cooperators to *these* fruits of cooperation is, as I said above, an essential component of the defence of reciprocity as elaborated from intuitions of fairness. Rawls puts aside the question of how to deal with the candidates most likely to be proposed as non-contributors, namely those with low levels of "physical capacities and skills" (see *PL*, pp. 183-6; cf. *Collected Papers*, p. 368, and G.A. Cohen, *Self-Ownership, Freedom, and Equality*, p. 225, n. 33). But elsewhere he offers us the basis of an answer to the worry when he claims that "[t]here is no question of determining anyone's contribution to society, or how much better off each is than they would have been had they not belonged to it" (*PL*, p. 279). The fruits of cooperation, even the material fruits of cooperation, are the product of much more than the work of those in easily identified manufacturing careers and the like. A person need not be in such a career to be properly counted a contributor. So the *real* line beneath which individuals count as non-cooperators will be much lower than the capacities of the normal subjects of these discussions would suggest. It might not be so implausible to suppose that individuals beneath that line are not owed a difference-principle-regulated share as a duty of *justice*. (C) | 9

6. Conclusion

The aim of this chapter was to present an argument for the egalitarianism of Rawls's principles which could slot into the hybrid contractualist framework that I have been recommending. The nature of that framework meant that the criteria for the argument's success were slightly different from those that one might ordinarily expect. For not only was it necessary to present a cogent argument from premises to the egalitarian conclusion, but also it was important that the intuitions of fairness with which the argument begins should not be so extreme that it would be utterly implausible to attribute them to each person in the Rawlsian society or, given the indoctrination objections I raised in the foregoing chapter, ours.

It does not seem unreasonable to suppose that we share the object conception of society as a fair system of cooperation between citizens regarded as free and equal. Elements of this conception are appealed to regularly in political argument. I think that it is also reasonable to suppose that we share the conception of fairness which ultimately determines the egalitarianism of the difference principle. To some extent this is an empirical matter: but to some extent too it is a matter of drawing out the implications of that conception of fairness and connecting it with other familiar and well-supported ideas, such the idea that the distribution of talents is a 'natural lottery', on the way to the egalitarian conclusion. I have been less quick to move from fairness to equality than others.²⁶⁸ Rawls too writes of *our* acceptance of the conditions on the original position, a device for representing the ideas of fairness involved in selecting the difference principle, as fair; clearly, he supposes that his elaboration of the concept of fairness will find support in the audience to

²⁶⁸ Temkin, for example. See the references to his work above.

which he addresses his theorising.²⁶⁵ If it seems reasonable to suppose that we share that conception, it cannot be less reasonable to suppose that individuals in a Rawlsian society, who have grown up and been educated under justice as fairness, would also share it.

The twin dangers that I warned of at the start of this chapter were: on one side, too strong a conception of fairness to be plausibly attributed to each person in a Rawlsian society; on the other, too weak a conception of fairness to support the egalitarian conclusions we wanted. The argument of this chapter addressed these dangers in different ways. The second was addressed by showing how the egalitarianism of the difference principle is substantial and well-supported by considering the arbitrariness of individuals' suitedness to a given cooperative system together with the consideration that the contribution of everyone is necessary to the creation of that system's specific social product. In other words, it is addressed by articulating the informal argument and making explicit its premises. I tried to address the first danger, meanwhile, by showing that motives of fairness associated with these premises, and not stronger motives of impartiality, underlie individuals' acceptance of the difference principle. Given our tendency to think that our duties of justice are primarily to our fellow-citizens, and not to all people, the motives required for an acceptance of reciprocity is much more plausibly attributed to each of us than those required for acceptance of impartiality.

Obviously this doesn't show that each person does actually endorse the conception of fairness that the informal argument draws on even if she does share the object conception. On a hybrid contractualist view, like any actual contractualist view, what is just is hostage to empirical facts about what people actually accept.

²⁶⁵ See e.g. *J-4F*, p. 80. Cf. *Theory*, pp. 580-2/508-10.

After we have set out the foundations, a great deal of argument is, as we have seen over this and the preceding chapter, to some extent speculative. The best we can do is hope to persuade the reader. But no less is true of the views of Locke, Hobbes, and Nozick. And we have the advantage that the reader's own views are a good guide to those of the individuals in whom we are interested. Putting the argument for the difference principle to her is a viable means of persuading her that others actually accept it in the sense we want.

Chapter 6: Rawlsian Contractualism, Actual Contractualism, and Left-Libertarianism

1. The hybrid view: an overview	246
2. The hybrid view and the actual contractualist tradition	253
3. The hybrid view and left-libertarianism	258
4. Conclusion	270

1. *The hybrid view: an overview*

The first two chapters of this thesis were concerned to describe two forms of response to the ideal of individualistic justification: what I called actual contractualism and what I called modal contractualism. In particular, I sought to distinguish certain structural characteristics of the two views. Actual contractualists, I argued, are marked by two essential features: the specification of a moral background and the specification of a legitimising process of unanimous acceptance. A key point about this legitimising process is that it involves an appeal to an actual exercise of each individual's agency in order to justify the imposition of new norms. Most actual contractualists also give an account of the moral conception so legitimised. Among these, 'dual condition' and 'mixed condition' actual contractualist theories argue that at least aspects of the moral conception that they advocate *ought*, for moral reasons independent of the actual contractualist process of legitimisation, to be legitimised via that process. And among *these*, 'unified' theories claim that the (relevant aspects of the) moral conception that they advocate is legitimised precisely because those individuals whose actual acceptance is

necessary accept (those aspects of) that moral conception for those moral reasons.

There is, I suggested, something very satisfying about unified dual condition actual contractualism. I then discussed the actual contractualist views of Hobbes, Locke, and Nozick in the context provided by this analysis, arguing that the emptiness of Hobbes's moral background in particular gives his view some of the attraction of a unified dual condition view.

Actual contractualism has a certain appeal as an interpretation of the individualistic justificatory ideal in that it requires real unanimity in individuals' actual views. It realises the ideal by requiring that the basis for proposed moral conceptions be actually accepted by each individual. Unanimity is something that even modal contractualists see as an ideal of individualistic justification; as we have seen, they sometimes suggest indeed that modal contractualism is appealing itself because it is the closest we can realistically get to it. However that may be, modal contractualism also manifests the following attraction which actual contractualism does not: it is concerned from the start that moral conceptions should be *objectively* justifiable, not just justifiable in the (even considered) opinion of each person.

Modal contractualism, then, has its own attractions independently of the fact that it purports to interpret the individualistic ideal. In Chapter 2 I argued that what makes it also an interpretation of that ideal is its identification of the Individualist Restriction on reasons for the rejection of proposed moral conceptions. This restriction requires that any such reasons appeal to a proposed conception's implications for single individuals, and not to implications for aggregations of individuals or to implications for the achievement of impersonal values. Therefore if the actual contractualist reason for rejection of a proposed moral conception—namely that some individual does not accept it—conforms to the Individualist

Restriction, we might hold out hope of combining the competing attractions of modal and actual contractualism into a single 'hybrid' view.

And this is indeed the core of the view I have been developing: that there is a modal contractualist case to be made for the necessary condition of justice employed by actual contractualist theories. That condition is a condition of actual acceptance, and we can discover a modal contractualist case for the construal of it that I argued for at the end of Chapter 3 in the Rawlsian idea of the burdens of judgment, according to which individuals can and do faultlessly differ in their sincere assessments of the moral and ethical facts. Because of the burdens of judgment, it is unreasonable to expect a person to accept a moral conception based on a moral outlook that she does not *actually* endorse.

This hybrid contractualist view is potentially very attractive. Both actual contractualism and modal contractualism have, as we have seen, strong appeal as interpretations of the ideal of individualistic justification, and the hybrid view promises to combine them both—even, indeed, to realise them more fully than either of the two approaches does on its own. But there are certain difficulties. The appeal of actual contractualism is compromised if its scope—the constituency of those whose actual acceptance is necessary for the justice of a moral conception—is restricted in ways which don't follow from the basis of requiring actual acceptance in the first place. But restrictions beyond those which follow from the burdens-of-judgment basis for actual contractualism appear to be justified by the modal contractualist framework which underlies the condition of actual acceptance. For modal contractualists have no reason to be concerned with the actual acceptance or rejection of a proposed moral conception by those who don't acknowledge, further, the modal contractualist reason that they have to find and comply with principles

that each person can reasonably be expected to accept. I called this the puzzle of scope reconciliation.

Moreover, among those who do acknowledge this modal contractualist reason we face a further difficulty which I called the arbitrariness puzzle: if they prioritise this reason, as it seems that they should, and what someone can reasonably be expected to accept is just what she *does* accept, then why wouldn't each person be ready to accept *anything* in order to reach agreement and so have complied with the reason that she has to find and comply with principles that each person actually accepts? Why isn't all we need for consensus mere coordination?

Drawing on Rawls's work, I argued in Chapter 3 that we can meet these two difficulties by affirming a particular conception of the role that the modal contractualist reason should play in the reasoning of those who acknowledge it. Acknowledging it should entail a readiness to reconsider and perhaps modify one's judgments about the moral and ethical facts. But a readiness to reconsider does not entail a willingness to abandon one's judgments in the face of others' rejection of them. It creates a tendency towards consensus, but a consensus determined by the range of (non-arbitrary) starting points rather than by arbitrary coordination.

This solves the arbitrariness puzzle. It also allows us to sidestep the puzzle of scope reconciliation. For what follows from the fact that someone might reasonably fail to change her view at all even though she acknowledges the modal contractualist reason and incorporates it into her reasoning is this. We—the theorists—must acknowledge the possibility that someone who *fails* to acknowledge the modal contractualist reason, and who therefore excludes herself from the modal contractualist's scope, might, even if she were reasonable, continue to accept and reject precisely the same moral conceptions that she accepts and rejects now. We therefore cannot ignore her acceptance and rejection, since it is possible that she

could reasonably reject the moral conception we propose. So we have reason to seek the acceptance of even those who are *prima facie* excluded from the modal contractualist's scope. Thus the restriction in the hybrid theory's scope which threatened to undermine its ability to share the attraction of straightforwardly actual contractualist theories is not, after all, effected.

Having thus distinguished the framework of a hybrid contractualist theory, I turned to an examination of Rawls's later views in order to fill this framework out into a full egalitarian view. I started by considering the strategy of political constructivism as a way to argue that individuals in a pluralistic society such as that regulated by justice as fairness, the moral conception ultimately advocated by Rawls, actually accept that moral conception. I noted that asking about actual acceptance in the Rawlsian society and no other (the 'head-first' approach) expresses our own acceptance of a dual condition actual contractualism: we suppose that justice as fairness is independently well-grounded, but that to be just it must also pass the actual contractualist test set by the hybrid theory. In ignoring other moral conceptions that might be accepted we manifest our supposition that these would not pass the regulative conception test that is part of all dual condition views. I also noted that the head-first approach brings with it the advantage of our familiarity with the normative views that are likely to be held by individuals in the Rawlsian society together with the disadvantage that it must respond to accusations that the actual contractualist test is met only through indoctrination.

Political constructivism relies on the 'shallowness' of the foundations we can give to justice as fairness. Calling them shallow expresses the fact that they are implicit in an 'object conception' of society as a fair system of cooperation between reasonable individuals who have their own conceptions of the good to pursue, rather than in the deeper fundamental judgments that we—the theorists—endorse

after our own consideration of the moral landscape. This fact makes it plausible to suppose that they might be shared by each person. More importantly, though, it creates room for us to suggest that they might be the focus of an 'overlapping consensus' of different comprehensive doctrines. In an overlapping consensus, each person affirms the foundations of justice as fairness and gives the values that the political constructivist elaborates from them normative priority over the other parts of her comprehensive view. Why should we agree that such a consensus would arise? I considered a number of reasons: because we suppose that people in the Rawlsian society will conceive of society in the way that permits the political constructivist presentation of justice as fairness; because we suppose that people are in general reasonable and so inclined to reconsider views which conflict with each other, which creates a tendency to consensus on shared elements; because of the fact that the values associated with our life together in society are so important, given the idea that society is the 'groundwork of our existence'; and because, as I went on to argue in Chapter 5, the concept of fairness involved in the object conception, and which is the basis of justice as fairness's egalitarianism, is more plausibly attributed to each person than the egalitarianism it produces might lead one to expect. These do not constitute proofs that the overlapping consensus will be secured, but they give us grounds for hope. This is not obviously a fault with the Rawlsian hybrid view: all actual contractualist theories rely to some extent upon speculation and predictions rather than proof. Good grounds for the predictions are the most we can ask.

The head-first approach provokes objections that justice as fairness can pass the actual contractualist test only through indoctrination, as I mentioned a moment ago. The remainder of Chapter 4 analysed and replied to this objection. I identified four forms that it might take and argued that each one fails. The hybrid view does not

rely on anyone's having false beliefs, or on anyone's failing to think through their beliefs. For the best reasons we can think of are offered in favour of justice as fairness as a regulative conception, and a society of justice as fairness meets the condition of full publicity whereby these same reasons are made publicly available. Moreover, the hybrid view does not rely on anyone's acting against their own interests. Not by the standards of justice as fairness, as the conception of primary goods as citizens' needs gives us reason to suppose that a society of justice as fairness would not be contrary to citizens' interests. And not by the standards of individuals' own diverse conceptions of the good, precisely because if the overlapping consensus is secured, then individuals have reasons stemming from their own conceptions of the good to support justice as fairness. Finally, the hybrid view does not rely on anyone's having been brought up in conditions other than those in which they should have been brought up. For, in the first place, the conditions in which individuals in the Rawlsian society are brought up are as close to ideal conditions of non-indoctrination as an objector concerned about indoctrination could hope for. And in the second place, the conditions under which justice as fairness is realisable can be reached without recourse to the sort of systematic indoctrination that would lead us to reject a theory if it were the only means of bringing the recommended moral conception about.

The final question that I addressed in Chapter 5 was: can we use the strategy of political constructivism to construct an egalitarian theory? That, I take it, is also part of Rawls's aim (though not initially in those terms) throughout his work. Consequently I attempted to reconstruct Rawls's argument for the difference principle as expressive of 'reciprocity' in order to show that the assumptions on which it is based could indeed be employed in a political constructivist presentation. This means that it would be plausible to see them as elaborations of elements in the

object conception that each person shares. A good test of this is that you and I, citizens in liberal Western democracies and among the best real-life counterparts to the imagined Rawlsian citizen, find the argument convincing. I tried, therefore, simply to bring out the moral intuitions—the unfairness of anyone's having a claim upon a greater share of the fruits of cooperation just in virtue of her suitedness to the cooperative system in question, given the necessity of the contribution of each person to the production of those fruits—upon which the argument is based and to explain how they support the egalitarianism of the difference principle. The hope is that you, the reader, will find this plausible. If so, and we can successfully make the same case to others, then egalitarian hybrid contractualist justice is achievable.

Much more would need to be said for a complete justification of Rawlsian liberalism by appeal to the hybrid theory. In particular, I have not addressed Rawls's first principle at all, and I have not argued for the primary goods metric beyond explaining how it must be based on 'a partial similarity in the structure of citizens' permissible conceptions of the good'. That is compatible with non-resourcist metrics, though Rawls himself rejects these. My aim has been only to show how egalitarianism, on whatever metric, can be justified on a hybrid contractualist view.

2. *The hybrid view and the actual contractualist tradition*

The account I have been developing offers a (dual, unified) hybrid contractualist condition for the justice of proposed moral conceptions and proposes that an egalitarian moral conception can meet it. Much of Chapters 4 and 5 have been devoted to explaining how these two parts are connected and how the just society

according to this account is connected with our own. This gives the account perhaps a peculiarly empirical aspect, considered next to many other contemporary theories of justice. Few such theories concern themselves with what each member of society actually thinks about *moral* matters except indirectly in the sense that at some point in a democratic society any policy must win the support of sufficient numbers if it is to be implemented. Of course they, like the argument of Chapter 5, often start from intuitions that the reader is presumed to share, and in that sense, since the theories are addressed to anyone who reads them, they aim to show that each person accepts their conclusions. But they don't explicitly ask about the likelihood of the assent of each person besides the reader. And in particular they don't explicitly employ such assent as a criterion of the justice of implementing their proposals. Even actual contractualist theories of justice tend not to be thought of as discussing acceptance in terms of proposed moral conceptions' fit with individuals' *moral* views. Rather, the proposed moral conceptions are said to meet the actual contractualist condition because of the likelihood of consent for non-moral mutually disinterested reasons by each person.

The hybrid contractualist account, then, asks about the likelihood of agreement on moral foundations among people who in many respects disagree fundamentally. Furthermore, it asks us to speculate about the likely development of individuals' moral views under particular historical conditions. This may seem to make it in some ways unphilosophical. To a certain extent this impression is undermined by drawing attention to the ways in which an ordinary philosophical argument from certain intuitions for the egalitarianism of Rawls's difference principle, doubles up, in addressing itself to you—a reader in the kind of society in which acceptance of that argument is hypothesised—as evidence for the empirical claim about the likelihood of acceptance to the extent that it is successful. But to some extent the

empirical aspect is anyway in keeping with the actual contractualist tradition. As we saw in Chapter 1, both Hobbes and Locke speculate about how life would be and what people would think under conditions prior to anyone's actual acceptance in the sense that is necessary for civil society as they conceive it. Hobbes can be read as supposing either that each person would seek his own conservation or that each person would want a situation in which the general observation of the laws of nature was guaranteed; in either case, he would 'seek peace' the only way possible: through the creation of the Leviathan. Locke too can be read in two ways: either as proposing a need for security or as proposing acceptance of the norms of the law of nature as the basis of individuals' agreement to the social contract. The point of these speculations is to ground a judgment about what people did actually accept or would actually have accepted, since that actual acceptance is a necessary condition of the justice of the moral conceptions being proposed. The proposal is hostage to the facts about what people *really* did or *really* would have accepted in just the way that the Rawlsian proposal I have offered is hostage to the facts about whether we do actually accept the egalitarian elaboration of intuitions of fairness.

The hybrid theory has advantages over the Hobbesian and Lockean approaches, however, in the following respect. Their speculations make no attempt to connect what they suppose that individuals would have accepted with the moral psychology of humans brought up specifically in the state of nature. They draw instead on what they take to be immutable human nature. Although of course our best understanding of human nature must play a part in our expectations about what individuals would accept or not, it is surely implausible nevertheless to suppose that it is uninfluenced by the institutions under which they have been brought up. So speculations about what individuals do or don't accept are properly also informed by knowledge about such institutions. Here, the Rawlsian head-first approach is

preferable: first, because it takes this idea seriously, concerning itself with the effect upon individuals that an upbringing under institutions of justice as fairness would have; and second, because we have first-hand knowledge of the normative beliefs prevalent under institutions similar to those of justice as fairness. It is the worries about indoctrination that this gives rise to—not at issue when one bases one's speculations upon assumptions about immutable human nature—that explains the other empirical aspect of the view I have been developing. A plausible account of the conditions of individuals' acceptance of justice as fairness as historically possible via processes other than upbringing under its institutions gives us reason to grant that a form of indoctrination for self-perpetuation is not the sole source of that acceptance.

All of this gives us, I believe, one way of understanding Rawls's conception of political philosophy and specifically justice as fairness as 'realistically utopian'.²⁷⁰ Justice as fairness, he writes, "probes the limits of the realistically practicable, that is, how far in our world... a democratic regime can attain complete realization of its appropriate political values".²⁷¹ The utopianism of the hybrid approach lies in the starting point of a Rawlsian society: we ask about individuals' outlooks under institutions which are not those of our world as it stands. Part of the story about the realistic practicability of justice as fairness, meanwhile, is told by the discussions of its stability for the right reasons. But part of it must also be told by the account of how we get to utopia from where we are. In this sense, a theory might be realistically utopian only with respect to particular historical circumstances: Hobbes

²⁷⁰ See *JAF*, pp. 4-5, 13; cf. *PL*, pp. lxi-lxii.

²⁷¹ *JAF*, p. 13.

and Locke could be realistically utopian for people in the state of nature, but they are not realistically utopian for us, now.

Hobbes and Locke are in another way not realistically utopian at all, though, because a theory is realistically utopian only when it is willing to *probe the limits* of the practicable. To take human nature as a starting point without paying attention to its mutability, to the role of a moral conception as educator,²⁷² is to concentrate unduly on the 'realistic'. Although Rawls invokes Rousseau's 'men as they are, and institutions as they might be', it would not be wrong to say that realistic utopianism requires us to consider both men and institutions as they might be. For human nature as it is gives us scope to consider the different ways in which we can be moulded.

The Rawlsian hybrid view differs from Hobbes's and Locke's views as well as their contemporary descendants' not only in the ways I have just described but in its novel basis for the requirement of actual contractualist legitimacy in the first place. Underlying Locke's actual contractualism is a conception of individual autonomy as what I called 'autonomous living'. This explains the requirement of consent for the legitimacy of the imposition on individuals of a moral conception (consent preserves autonomous living); but as I argued, it also leads to a less thoroughgoing actual contractualist condition than we might hope. For the concern with autonomous living rules out a moral background in which the conditions of that autonomous living (as Locke conceives them) aren't secure. Otherwise consent can't *preserve* autonomous living. The moral background, of course, isn't subject to

²⁷² Rawls stresses this function of what I'm calling a moral conception at *PL*, p. 71. (Note that he is not referring to specifically educational policies associated with a given moral conception. See Chapter 4, section 7, note 218 above.)

the actual contractualist test. Hobbes's actual contractualism, meanwhile, can be read as much more thoroughgoing than Locke's in this sense: the moral background is normatively virtually empty. As a result *any* moral conception must pass the actual contractualist test in order to be justified. But it's not altogether clear what the ground of Hobbes's imposition of the test is in the first place; he simply claims that obligations, including the obligation not to exceed the bounds of the restricted liberty that individuals have under the sovereign, can arise only through voluntary acts. The idea of the burdens of judgment leads to an actual contractualist test which has a rationale as clear as Locke's but which can consistently be as thoroughgoing as Hobbes's. If it's reasonable to differ in moral outlooks and so unreasonable to expect others to accept a moral outlook that one actually accepts oneself but that they do not, then we have reason, according to the modal contractualist framework that I discussed in Chapters 3 and 4, to seek and implement a moral conception based on fundamental elements which no one actually rejects. This is an actual contractualist test, but its basis is new with Rawls.

3. The hybrid view and left-libertarianism

I said at the outset of this thesis that my aim was to distinguish a way of combining egalitarianism with the attractions of actual contractualism. My attempt to do so is not unique. In recent years there has been a revival of interest in what has come to be known as 'left-libertarianism' which, as its name suggests, combines the actual contractualism implicit in libertarian theories such as Nozick's with the

egalitarianism of the left.²⁷³ I want to end with a brief comparison of this approach with my own and to say why I take the hybrid Rawlsian view to be preferable. My fundamental point is that left-libertarianism, though preferable to right-libertarianism from an egalitarian point of view, is nevertheless no better from an actual contractualist point of view. The hybrid view, on the other hand, is preferable in both respects.

Left-libertarian views involve two essential commitments: individual self-ownership and egalitarian world-ownership. The commitment to individual self-ownership is reflected in the assertion of a set of Lockean natural rights which, among other things, create an actual contractualist test for the imposition of any moral conception beyond that entailed by the natural rights themselves. Left-libertarians are typically political voluntarists.²⁷⁴ For the imposition of a moral conception involves coercion beyond that involved in the maintenance of our natural rights,²⁷⁵ and such coercion infringes individual self-ownership in the sense

²⁷³ For a summary of the left-libertarian position see Peter Vallentyne, "Left-Libertarianism: A Primer". For more detailed defence see Hillel Steiner, *An Essay on Rights*; Michael Otsuka, *Libertarianism without Inequality*; and Vallentyne, Steiner, and Otsuka, "Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant: a Reply to Fried." For criticism see Barbara Fried, "Left-Libertarianism: A Review Essay" and "Left-Libertarianism, Once More: A Rejoinder to Vallentyne, Steiner, and Otsuka"; and Mathias Risse, "Does Left-Libertarianism Have Coherent Foundations?"

²⁷⁴ See for example Otsuka, *Libertarianism without Inequality*, Chapter 5.

²⁷⁵ Lockean left-libertarians suppose that even the maintenance of our natural rights requires each person's actual acceptance if it is performed by a coercive state. But that is because a coercive state derives its status as such from the delegation to it of each person's natural right to punish. In delegating her natural right to punish in this way each person relinquishes it. After delegating it, she may not exercise it herself, and may be coerced not to exercise it. Therefore state coercion even

affirmed by the left-libertarian.²⁷⁶ The commitment to egalitarian world-ownership, meanwhile, is reflected in the assertion of egalitarian conditions on the permissible use or appropriation of external (to the person) natural resources.²⁷⁷ (These conditions are related to the original ownership of those resources: jointly owned, owned in common, or unowned.) So, for example, ‘Georgist libertarians’ hold that “agents must pay the full competitive value of the natural resources that they appropriate” with the payment being shared in an egalitarian manner among the remaining agents.²⁷⁸ Others argue that agents may appropriate natural resources as long as no one appropriates a share that is more than is compatible with each person’s having an equally advantageous share, where ‘equally advantageous’ is understood in luck egalitarian terms.²⁷⁹ Since I am interested in left-libertarianism as a competitor to the hybrid view, I shall restrict my focus to those left-libertarians

when it is restricted to the maintenance of our natural rights in the way that the Lockean imagines is coercion which is in one sense beyond that is involved in the maintenance of our natural rights.

²⁷⁶ Left-libertarian rights of self-ownership must be strong enough to give plausible substance to the idea that they articulate a conception of individuals as self-owners, but not so strong as to eliminate the possibility of all action given the at least small chance inherent in any action of unforeseen infringement of such rights (see Otsuka, *Libertarianism without Inequality*, pp. 13-15).

²⁷⁷ See Vallentyne, “Left-Libertarianism: A Primer”, §4. ‘External (to the person)’ admits in left-libertarian thought of a wide range of interpretations, not all of which are naturally included under that description. Steiner, for example, effectively includes germ-line genetic information as external in the relevant sense.

²⁷⁸ Vallentyne, “Left-Libertarianism: A Primer”, p.11. This is the view held by Steiner, among others.

²⁷⁹ This is Otsuka’s view. See *Libertarianism without Inequality*, pp. 22-35.

who “wish to demonstrate that their libertarian commitments are at least nearly fully consistent with the egalitarianism of people such as Rawls and Dworkin.”²⁸⁰

Right-libertarian views such as Nozick’s can be seen as having two core components. The first is a commitment to rights of self-ownership, and as G.A. Cohen has argued, this is the source of much of libertarianism’s attraction. The second is a commitment to—at most—only minimal distributive constraints on world-ownership. In Nozick’s presentation of libertarianism in *Anarchy, State, and Utopia*, both of these commitments are elaborated simply as consequences of the natural rights that Nozick takes each person to have in virtue of their inviolable moral status. The distinctive left-libertarian move is to see that affirming rights of self-ownership need imply nothing with respect to world-ownership, and hence that the attraction of self-ownership is compatible with egalitarian distributive constraints on world-ownership.²⁸¹ Hence the twin left-libertarian commitments.

My actual contractualist analysis of libertarianism in Chapter 1 suggests, however, that another source of libertarianism’s attraction might be the fact that the (coercive) imposition of any distributive constraints in a Nozickean world would have to be subject to an actual contractualist test. Nozick’s moral background is broadly *laissez-faire* with respect to world-ownership: more or less any distribution of worldly resources is compatible with it. Moreover, no particular distribution would be the result, he thinks, of individuals’ acts of acceptance against that moral

²⁸⁰ Vallentyne, Steiner, and Otsuka, “Why Left-Libertarianism is not Incoherent, Indeterminate, or Irrelevant”, p. 211.

²⁸¹ Otsuka plausibly suggests that in fact self-ownership maintains its attraction only if it implies sufficient world-ownership rights to prevent anyone’s being forced by necessity to work for anyone else. But this, in his view, is (at least “contingently”), reconcilable with an egalitarian distribution of worldly resources. See *Libertarianism without Inequality*, p. 31-4.

background, since in his view no ‘patterned’ or ‘end-state’ distributive principles would be among those accepted. But his argument nevertheless leaves open the possibility that as a result of its voluntary acceptance (which could be in principle irreversible) a patterned or end-state distributive principle could be coercively imposed on individuals’ holdings. It would reduce the attraction of his view (and demonstrate him to be less of an actual contractualist than I have supposed him to be) if the imposition of such principles was prevented by the inalienability of the rights of world-ownership which form the moral background rather than merely by the principles’ non-acceptance.

Some left-libertarians—those who see aspects of what right-libertarians would see as part of the self as instead part of the range of worldly resources—don’t endorse a right of *self*-ownership in any form that right-libertarians would recognise.²⁸² Theirs can hardly then be said to be a reconciliation of the actual contractualism of libertarianism and egalitarianism, since the actual contractualism of libertarianism is not secured at all. Those who are sufficiently talented in a scheme which includes talents in an egalitarian division of external resources may, for example, find themselves involuntarily yet justly coerced to benefit others (by employing their talents as the upshot of their egalitarian division dictates) in a way that clearly violates the principles of self-ownership affirmed by right-libertarians.

But others endorse a conception of the self which doesn’t so obviously conflict with the actual contractualist right. Yet even these may fail to secure the actual contractualist attraction of Nozick’s view. As we have seen, left-libertarians are committed to an egalitarian moral background where Nozick’s is *laissez-faire*. But

²⁸² For example: Steiner treats individuals’ germ-line genes as a natural resource and therefore subject to the payment of rent, as more obviously external natural resources are.

insofar as the distributive principles that this involves may not be overridden by individuals' unanimous actual acceptance, left-libertarians fail to secure much of the actual contractualist attraction of Nozick's view. What they gain in egalitarianism, that is, they lose in actual contractualism. This is not to say that left-libertarians *are* committed to immutable egalitarian distributive principles. They may or may not be. But to the extent that they are, they impose those principles at the expense of some of the actual contractualist attractions of libertarianism.

Insofar as the left-libertarians' egalitarian moral background may be overridden by unanimous acceptance of some other distributive conception, on the other hand, they are clearly no worse off in terms of actual contractualist allure than Nozick is. But it remains the case that they affirm a distributive conception as part of the moral background, on the same footing as the rights against being coerced in accordance with any moral conception that isn't actually accepted. This conception is consequently not subject to the actual contractualist test: the left-libertarian's egalitarian principles are, like her libertarian principles, taken to be coercively enforceable even before anyone has agreed to anything. (This is why libertarian views are global rather than national in scope: they operate at a level more fundamental than that of the voluntary associations that they see as political societies.) But coercive enforcement is, as we've seen, one prime candidate for individualistic justification; exempting egalitarian principles from such justification by excluding them from the actual contractualist test diminishes the attraction created by incorporating it. The same is true, of course, of the libertarian principles of self-ownership. But in the first place, the egalitarian principles in particular seem to necessitate the institutions of a global state—no set of feasible rules for individuals, however successfully followed, could conceivably ensure the attainment

of egalitarian justice on a left-libertarian view such as Otsuka's or Vallentyne's²⁸³—about which it is then natural to ask why *that* is not subject to the actual contractualist test when any sub-states are.²⁸⁴ In the second place, the libertarian principles are much more intuitively understood as *part* of the view's actual contractualism than are the egalitarian principles. And in the third place, increasing the range of principles subject to the actual contractualist test is, from the actual contractualist point of view, obviously preferable to decreasing them. Since actual contractualism has a certain attraction derived from the fundamental attraction of

²⁸³ Cf. Rawls on 'background justice' and the necessity of a basic structure at *PL*, pp. 265-9 and *J.4E*, pp. 52-5.

²⁸⁴ In *Libertarianism without Inequality*, Otsuka admits that there is "a case for voluntarism...at the interpolitical [i.e. global] level" (p. 109) but claims that it is overridden by the necessity of avoiding "the disorder and chaos which would ensue in the absence of such a governing body" (p. 108). He therefore supposes that an interpolitical governing body is justified. It is hard to see how, if this necessity does not override the case for voluntarism at the subglobal level (suppose that no one consents to give up her rights to punish), it should be able to do so at the global level. (In fact, for Otsuka, it *does* override the case for voluntarism at the subglobal level in many instances: if the population is sufficiently large that individuals have no choice but to "live, intermingle, and interact with others within the confines of a community that is much larger or more complex than a hamlet" [p. 94] then individuals have "an obligation to relinquish these rights to legislate and punish and instead to place themselves under a competent and effective common government" [pp. 93-4]. Otsuka counts himself a voluntarist anyway because he supposes that individual consent sufficient for voluntarism is given even in these situations if the range of possible communities that it is open to someone to join includes some that are sufficiently attractive for the relevant individuals [see pp. 103-7]. But this being so, one might wonder whether by his own lights, at any rate, Otsuka is an interpolitical voluntarist too. [The interpolitical government is after all attractive, given the disorder and chaos that would obtain without it.] For most libertarians, I take it that the interpolitical government would plainly *not* be voluntaristic.)

the ideal of individualistic justification (see Chapter 1), maximising the range of principles subject to its test maximises a theory's attraction in that sense.

So while left-libertarians are no *worse* off from an actual contractualist point of view than right-libertarians, they are no *better* off either. Left-libertarianism, that is, represents an improvement on right-libertarianism in *one* dimension (egalitarianism) but not in another. The hybrid view, by contrast, represents an improvement in both. The strong egalitarianism of justice as fairness enters the Rawlsian hybrid view at a less fundamental level than it enters the left-libertarian view. Given a set of cooperating individuals each of whom is taken to bear ~~a~~ comprehensive actual contractualist right against coercion ~~to~~ to behave in accordance with a moral conception that she does not accept, we ask: can egalitarianism pass the actual contractualist test?

This route is, of course, one that left-libertarians could take. They could affirm the libertarian rights of self-ownership but make no claims whatsoever about the justice of any distributive moral conception. One justification that Michael Otsuka offers for his commitment to the egalitarian moral conception which forms part of his left-libertarian moral background is that it reflects intuitions about fairness.²⁸⁵ As we saw in Chapter 5, this is also, indirectly, the source of the hybrid Rawlsian view's commitment to egalitarianism. That commitment reflects intuitions of fairness concerning the distribution among cooperators of the fruits of their cooperation, given the arbitrariness of each person's suitedness to that particular scheme. (These intuitions reflect a slightly weaker conception of fairness than do Otsuka's—they imply reciprocity rather than impartiality and so domestic egalitarianism rather than global egalitarianism—but that, I take it, is an advantage

²⁸⁵ See *Libertarianism without Inequality*, p. 23.

when we try to argue that each person actually does accept egalitarianism.) If Otsuka is convinced that the intuitions of fairness to which he appeals are shared by his audience, or would be shared if the audience thought carefully enough about their own views, a libertarian version of the hybrid view (where all enforceable rules of property in and distribution of external resources must be universally actually accepted) might represent a way to increase the attraction of his reconciliation of libertarianism and equality. (Indeed, if we had accepted Rawls's supposition that the basic liberties follow directly from acknowledgment of the burdens of judgment [see Chapter 3 above], we might have ended up with a view not dissimilar to this.)

The resultant view would be preferable, then, because in one way it would have become more thoroughgoing from an actual contractualist point of view. The major differences remaining between the hybrid theory and the revised left-libertarian view would then be twofold. First and foremost, the hybrid theory involves a moral background which is normatively empty whereas the revised left-libertarian view would still include libertarian rights of self-ownership as part of that background even once the egalitarian principles had been shifted to be subject to the actual contractualist test. I have already suggested that it maximises the hybrid theory's actual contractualist attraction that its moral background is empty; and indeed, this poses the following challenge to left-libertarians which is connected to a further objection offered by Barbara Fried in her review essay. One important basis for the left-libertarian affirmation of the libertarian right of self-ownership is its "anti-paternalistic and anti-moralistic implications".²⁸⁶ Nozick justifies rights of self-ownership by appeal to "the fact that there are distinct individuals each with his *own*

²⁸⁶ *Libertarianism without Inequality*, p. 2.

life to lead”,²⁸⁷ making it clear that what’s important about *that* is that the meaning of life derives from “[a] person’s shaping his life in accordance with some overall plan” that he has chosen.²⁸⁸ Certainly this conception of autonomy seems to justify some kind of actual contractualist right. But one might wonder whether specifically the libertarian right of self-ownership is the only way to do the idea justice. After all, the libertarian right of self-ownership prevents some people shaping their lives in accordance with some overall plans that they might choose—not just in the sense that it is compatible with a person’s starving if she is sufficiently unfortunate or careless, but in the sense that it rules out anyone’s living any life that involves violating others’ rights of self-ownership. As Fried notes, invoking Hohfeld and Coase, any distribution of rights and duties involves arbitrating between conflicting interests.²⁸⁹ Arbitrating between them therefore means privileging some over others. But the point about these conflicting interests is that they all represent individuals’ ways of giving meaning to their lives, so arbitrating between them inevitably means privileging certain ways of doing this over others. There will be assignments of rights and duties other than those affirmed by libertarians and which privilege different ways for individuals to give meaning to their lives. These could equally be justified by appeal to the very values invoked by libertarians. Why, then, should we choose specifically libertarian rights of self-ownership as the moral background?

²⁸⁷ *ASU*, p. 34.

²⁸⁸ *ASU*, p. 50. More precisely, Nozick justifies humans’ moral status by appeal to that fact. But that moral status is taken to explain our rights of self-ownership. It would be strange if there were no explanatory connection from the grounds of that moral status to the kinds of rights that beings with such status are supposed to have.

²⁸⁹ See “Left-Libertarianism: A Review Essay”, pp. 72ff.

In this context it seems to me that the hybrid theory's normatively empty moral background (i.e. involving only the right against coercion to conform to principles that one does not actually accept) offers an appealing way to negotiate the problems posed by the indeterminacy in appropriate rights allocation which follows from affirmation of the fundamental basis that the Nozickean libertarian adopts. If any particular allocation privileges the interests of some (as determined by that fundamental basis) over those of others, surely a radically individualistic view should be unwilling to pick one without universal acceptance of it among those whose conduct it will regulate? Of course, in the absence of any particular allocation it is no less true that the interests of some will be served better than those of others. But the theorist refuses to grant this situation and the bargaining positions it creates the endorsement of morality. We have neither justice nor injustice without the imposition of some moral conception, and none is imposed. The view is, moreover, more thoroughgoing yet than even the revised libertarian view. It subjects everything to the actual contractualist test.

The second major difference is that the hybrid theory adopts the 'head first' approach. But given the left-libertarian's own commitment to egalitarianism and the advantages of this approach that I described in Chapter 5, I see no reason why the left-libertarian should object to this. So long as one can avoid the various suspicions of indoctrination that I tried to deal with in Chapter 4, there seems to be no great advantage for one with egalitarian intuitions of fairness in speculating about the kinds of reasons that individuals prior to the establishment of any cooperation according to egalitarian principles might have taken themselves to have. (And next to the other alternative of imposing an egalitarian moral conception as part of the moral background, it does not seem objectionably tendentious.) Recall that considering the acceptance-grounding views of individuals who have been shaped

by a given moral conception is not the same as taking that moral conception as the moral background for the actual contractualist test (see Chapter 4, section 7 above). The head-first approach does not alter the moral baseline for bargaining positions in an actual contractualist view; rather, it determines the types of moral reasons that individuals are likely to take themselves to have. The baseline is set by the presence of the actual contractualist right and any rights of self-ownership whether we adopt the head-first approach or not. Of course, individuals *brought up* in the egalitarian world which the imposition of such a baseline would involve might be very unlikely to agree upon an egalitarian moral conception. But it isn't clear why we should care especially about what *they* would accept as long as we're not working with a conception of acceptance which is based on the self-interest of humans conceived as having immutable interests and nature. (In that case, caring about what *they* would accept is caring about what *anyone* would accept.) And that is not the conception of acceptance which libertarians of any stripe are wedded to: indeed, they are motivated in part by an appreciation that individuals' acceptance may properly be guided by their desire to live in a community which shares their own normative outlook.²⁹⁰ It is perfectly reasonable to suppose that those whose acceptance we care about are (a) those who are subject to the moral conception we are interested in defending, and (b) those whose societies pave the way for the just implementation of that conception and thus provide us with a defence against charges of indoctrination. Discussion of individuals whose views are shaped by experience of the state of nature, then, even if we could provide a credible

²⁹⁰ Consider for example Nozick's utopias (*ASU*, Chapter 10) or Otsuka's pluralistic political societies and monies (*Libertarianism without Inequality*, pp. 101-7).

foundation for it, is surely irrelevant. The hybrid contractualist view is superior in this respect as well.

4. Conclusion

That concludes my comparison of the hybrid view and left-libertarianism and this thesis as a whole. I hope now to have achieved the following aims. First, to distinguish two contrasting approaches to the contractualist ideal of individualistic justification and to suggest how they might be combined into one justificatory framework. Second, to explain—drawing on ideas that I find in Rawls's thought—how a moral conception might meet the 'hybrid' contractualist test of justice which this combined framework creates and to consider the ways in which a proponent might meet certain major objections to her approach. Third—again drawing on Rawls's thought—to show how the egalitarianism of justice as fairness would meet the contractualist condition of justice on the hybrid view. And, finally, fourth, to compare the resulting view with left-libertarianism, a competing approach which also aims to reconcile actual contractualism and egalitarianism. A thesis of this length is not sufficient to consider, much less resolve, every question that these aims give rise to. I have tried, however, to address the most important among them.

On the resultant view of egalitarian contractualist justice, the achievement of a just egalitarian society is more precarious than one might hope. Its justification appeals to the peculiarly empirical factors that I discussed in section 2 above. In that sense it may seem a less satisfying theory than others. I hold, however, that this is a reflection of the real problems that reconciling contractualism and egalitarianism

presents. A theory which seems to avoid them fails to appreciate the depth of the issue.

Bibliography

Barry, B. (1995). 'John Rawls and the Search for Stability', *Ethics*, 105 (4), pp. 874-915.

Berlin, I. (1990). 'On the Pursuit of the Ideal', in I. Berlin, *The Crooked Timber of Humanity*, H. Hardy (ed.). London: John Murray.

———. (2002). 'Two Concepts of Liberty', in I. Berlin, *Liberty*, H. Hardy (ed.). Oxford: Oxford University Press, pp. 166-217.

Blake, M. (2001). 'Distributive Justice, State Coercion, and Autonomy', *Philosophy & Public Affairs*, 30 (3), pp. 257-296.

Brink, D. (1993). 'The separateness of persons, distributive norms, and moral theory', in R.G. Frey and C.W. Morris (eds), *Value, Welfare, and Morality*. Cambridge: Cambridge University Press, pp. 252-289.

Cohen, G.A. (1995). *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press.

———. (1997). 'Where the Action Is: On the Site of Distributive Justice', *Philosophy & Public Affairs*, 26 (1), pp. 3-30.

———. (2007). *Rescuing Justice and Equality*. Manuscript.

Darwall, S. (2006). 'Contractualism, Root and Branch: A Review Essay', *Philosophy & Public Affairs*, 34 (2), pp. 193-214.

Dworkin, R. (1975). 'The Original Position', in N. Daniels (ed.), *Reading Rawls*. Stanford: Stanford University Press, pp. 16-52.

Frankfurt, H. (1971). 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68 (1), pp. 5-20.

Freeman, S. (2003). 'Congruence and the Good of Justice', in S. Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, pp. 277-315.

Fried, B. (2003). "If You Don't Like It, Leave It": The Problem of Exit in Social Contractarian Arguments', *Philosophy & Public Affairs*, 31 (1), pp. 40-70.

———. (2004). 'Left-Libertarianism: A Review Essay', *Philosophy & Public Affairs*, 32 (1), pp. 66-92.

———. (2005). 'Left-Libertarianism, Once More: A Rejoinder to Vallentyne, Steiner, and Otsuka', *Philosophy & Public Affairs*, 33 (2), pp. 216-222.

Garrison, J.W. (1986). 'The Paradox of Indoctrination: A Solution', *Synthese*, 68, pp. 261-273.

Gibbard, A. (1991). 'Constructing Justice', *Philosophy & Public Affairs*, 20 (3), pp. 264-279.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Hobbes, T. (1968). *Leviathan*, ed. C. B. MacPherson. London: Penguin Books.

Kelly, M., and McPherson, L. (2001). 'On Tolerating the Unreasonable', *The Journal of Political Philosophy*, 9 (1), pp. 38-55.

Locke, J. (1960). *Two Treatises of Government*, ed. Peter Laslett. Cambridge: Cambridge University Press.

Munoz-Dardé, V. (2007). Bound Together: Claims of Others and the Needs of Selves. Manuscript.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

———. (1979). 'Equality', in T. Nagel, *Mortal Questions*. Cambridge: Cambridge University Press, pp. 106-127.

———. (1991). *Equality and Partiality*. New York: Oxford University Press.

Narveson, J. (1988). *The Libertarian Idea*. Philadelphia: Temple University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Otsuka, M. (2003). *Libertarianism without Inequality*. Oxford: Oxford University Press.

———. (2006). 'Saving Lives, Moral Theory, and the Claims of Individuals', *Philosophy & Public Affairs*, 34 (2), pp. 109-135.

Parfit, D. (1991). 'Equality or Priority?'. *The Lindley Lecture, University of Kansas*, pp. 1-42

———. (2003). 'Justifiability to Each Person', *Ratio*, XVI, pp. 368-390.

Quong, J. (2004). 'The Rights of Unreasonable Citizens', *The Journal of Political Philosophy*, 12 (3), pp. 314-335.

Rawls, J. (1971). *A Theory of Justice* (revised edition 1999). Cambridge: Mass.: Harvard University Press.

———. (1993). *Political Liberalism*. New York: Columbia University Press.

———. (1999). *Collected Papers*. Cambridge, Mass.: Harvard University Press.

———. (2001). *Justice as Fairness: A Restatement*, E. Kelly (ed.). Cambridge, Mass.: Harvard University Press.

———. (2007). *Lectures on the History of Political Philosophy*, S. Freeman (ed.). Cambridge, Mass.: Harvard University Press.

Raz, J. (1986). *The Morality of Freedom*. Oxford: Clarendon Press.

———. (1990). 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy & Public Affairs*, 19 (1), pp. 3-46.

Risse, M. (2004). 'Does left-libertarianism have coherent foundations?', *Politics, Philosophy & Economics*, 3 (3), pp. 337-364.

Scanlon, T.M. (1975). 'Preference and Urgency', *The Journal of Philosophy*, 72 (19), pp. 655-669.

———. (1998). *What We Owe To Each Other*. Cambridge, Mass.: Harvard University Press.

———. (2003). ‘Rawls on Justification’, in S. Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, pp. 139-167.

———. (2003). ‘Replies’, *Ratio*, XVI, pp. 424-439.

———. (2003). ‘Contractualism and Utilitarianism’, in T.M. Scanlon, *The Difficulty of Tolerance*. Cambridge: Cambridge University Press, pp. 124-150.

Scheffler, S. (1994). ‘The Appeal of Political Liberalism’, *Ethics*, 105 (1), pp. 4-22.

———. (2006). ‘Is the Basic Structure Basic?’, in C. Sydnorowich (ed.), *The Egalitarian Conscience*. Oxford: Oxford University Press, pp. 102-129.

Simmons, A.J. (1979). *Moral Principles and Political Obligations*. Princeton: Princeton University Press.

———. (1999). ‘Justification and Legitimacy’, *Ethics*, 109 (4), pp. 739-771.

Steiner, H. (1994). *An Essay on Rights*. Cambridge, Mass.: Blackwell Publishing.

Temkin, L. (2000). ‘Equality, Priority, and the Levelling Down Objection’, in M. Clayton and A. Williams (eds), *The Ideal of Equality*. Basingstoke: Palgrave, pp. 126-161.

———. (2001). ‘Inequality: A Complex, Individualistic, and Comparative Notion’, *Philosophical Issues*, 11, pp. 327-353.

Vallentyne, P. (2000). ‘Left-Libertarianism: A Primer’, in P. Vallentyne and H. Steiner (eds), *Left-Libertarianism and Its Critics: The Contemporary Debate*. London: Palgrave, pp. 1-20.

Vallentyne, P., Steiner, H., and Otsuka, M. (2005). ‘Why Left-Libertarianism is Not Incoherent, Indeterminate, or Irrelevant: A Reply to Fried’, *Philosophy & Public Affairs*, 33 (2), pp. 201-215.

Walzer, M. (1981). 'Philosophy and Democracy', *Political Theory*, 9 (3), pp. 379-399.

Williams, B. (1995). 'The Point of View of the Universe', in B. Williams, *Making Sense of Humanity and other philosophical papers*. Cambridge: Cambridge University Press, pp. 153-171.